MOMENT EXPANSIONS FOR ROBUST STATISTICS

BY

KEAVEN MARTIN ANDERSON

TECHNICAL REPORT NO. 7
MARCH 12, 1982

U.S. ARMY RESEARCH OFFICE
RESEARCH TRIANGLE PARK, NORTH CAROLINA
CONTRACT NO. DAAG29-79-C-0166

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Moment Expansions for Robust Statistics

By

Keaven Martin Anderson

Technical Report No. 7
March 12, 1982

U.S. Army Research Office
Research Triangle Park, North Carolina
Contract No. DAAG29-79-C-0166

Department of Statistics
Stanford University
Stanford, California

Abstract

Our objective is to give asymptotic expansions for moments of standardized statistics based on n independent, identically distributed random variables as $n \to \infty$. The basic premise is that a simple tail condition on the underlying distribution which implies the moments of a standardized quantile converge to the moments of an appropriate normal distribution is sufficient to assure the validity of asymptotic moment expansions for many statistics which are resistant to outliers.

The primary result we present gives sufficient conditions for the validity of moment approximations based on moments of Taylor's series approximations which are obtained by using functional differentiation. We apply the theory to some L- and M-estimates and present a Monte Carlo study to show that the approximations for the variance of statistics based on small to moderate sample sizes can be quite good.

Prior to studying the above general problem we consider the problem of the convergence of the moments of a standardized quantile to those of an appropriate normal distribution. Our proof of moment convergence requires fewer non-tail conditions on the underlying distribution than were used in previously published results. We also extend the result to show necessary and sufficient tail conditions on the underlying distribution for convergence of the moment generating function of a standardized quantile to that of a normal distribution.

Acknowledgements

I would first like to express my gratitude to my adviser, Vernon Johns, who encouraged me to extend my work on quantiles to the work presented in this thesis. His knowledge has helped considerably in the development of the work presented, and his advice has helped to clarify many of the ideas I have attempted to express.

James Fill read this thesis very carefully and found many errors in earlier versions. His comments have improved many proofs as well as the clarity of exposition throughout this work. I would also like to thank Brad Efron for serving on my reading committee, and for sharing his insight into some of the problems considered here.

Barry Eynon independently came up with ideas similar to some presented in this thesis. I would like to thank him for 'sharing' the topic, and for sharing his ideas on it. I am grateful to James Reeds for discussions which have contributed much to my understanding of functional differentiation, a primary tool used in this work. I would also like to thank Lars Holst, who encouraged me to write up the initial result I obtained on moment generating functions of quantiles.

I am very grateful to Alice Whittemore, who has arranged for my financial support for much of the time I have been at Stanford. I would like to thank Paul Switzer, who provided funds for the computer account used to print this thesis, and Max Díaz, who helped resolve problems with the 'typesetting' of this work.

Finally, I would like to thank my parents and my wife, Kay, for the much needed and appreciated support they have given me.

Contents

List of Tables

List of Figures

Chapter 1

Summary and literature review

§1.1 Introduction.

This chapter summarizes the results obtained and gives a literature review for the two basic problems which we consider. First we consider the convergence of the moments of a standardized quantile to the moments of a normal distribution, and then we move on to summarize results on asymptotic expansions for moments of robust statistics.

§1.2 Standardized quantiles.

In chapter 2 we will attack the problem of the convergence of the moments of a standardized quantile to the moments of a normal distribution using direct methods; *i.e.*, we will write down integral expressions for expectations and use standard tools from analysis to obtain results. Although much of the theory of the rest of the thesis is not heavily dependent on this chapter, some basic ideas are illustrated without the additional tools and technical problems of later chapters.

For independent, identically distributed random variables with distribution function F, necessary and sufficient tail conditions needed for convergence of moments of standardized quantiles are

$$0 < \alpha_+ = \liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log x} \quad \text{and} \quad 0 < \alpha_- = \liminf_{x\to\infty} \frac{-\log F(-x)}{\log x}. \tag{1.2.1}$$

Blom (1958), Sen (1959) and Bickel (1967) have considered equivalent conditions, but have not explicitly defined quantities which are quite as useful as α_+ and α_-. We will discuss the relation of the values of α_+ and α_- to the existence of moments, the existence of moments of order statistics, regular variation, and hazard functions, as well as their relation to the convergence of moments of

standardized quantiles. The only other condition on F we require is that it be differentiable at the quantile of interest. Other authors have required that F be an absolutely continuous distribution, but we have developed a simple proof which does not require this condition.

Most of the results of chapter 2 will be concerned with the expectations of functions of standardized quantiles. This general approach will allow us to consider convergence in distribution and convergence of the moment generating function as well as convergence of moments. Whereas the result on convergence in distribution is contained in results given by Smirnov (1952) and Wretman (1978), and the results on convergence of moments are variations on previous results as discussed above and in chapter 2, the result on the convergence of the moment generating function is believed to be completely new.

§1.3 Robust statistics.

In chapter 3 we will extend the results of chapter 2 in two ways. First, we will show that the conditions of (1.2.1) are sufficient to assure convergence of moments of many statistics which have bounded influence functions. Second, we give higher order expansions of moments using functional differentiation.

Previous applications of functional differentiation in statistics have been proofs of versions of the central limit theorem, the theory of Edgeworth expansions, the law of the iterated logarithm, and the Berry-Esséen theorem. Serfling (1980), Reeds (1976), and Huber (1981) present surveys on the applications of functional differentiation in statistics. To our knowledge the theory has not been used to prove the validity of asymptotic moment expansions.

The theory presented here involves showing that for a functional statistic T, an underlying distribution F, and an empirical distribution function F_n an expansion of the form

$$\mathrm{E}[(T(F_n) - T(F))^r] = \mathrm{E}\left[\left(\sum_{j=1}^{k} T_j(F; F_n - F)/j!\right)^r\right] + o(n^{-(r+k-1)/2}) \tag{1.3.1}$$

is valid under some assumptions. We have used a version of Fréchet differentiation to prove this result. The condition of Fréchet differentiability on T is a strong one. If a functional T is Fréchet differentiable then the corresponding functional statistic $T(F_n)$ is, in general, resistant to outliers. It is this fact that allows us to use the same tail conditions which are used for quantiles to show convergence of moments of many other statistics. That Fréchet differentiability is a stronger condition than we might like is indicated by the fact that quantiles do not correspond to a Fréchet differentiable functional, and yet quantiles are statistics which are resistant to outliers whose moments converge under the tail conditions we use.

At the end of chapter 3 we give (previously known) results to aid in computing the right hand side of (1.3.1) to within $o(n^{-(r+k-1)/2})$. In chapter 4 we develop formulas for these approximations for M-estimates of location which are not scale invariant and for L-estimates. In particular, we give formulas for first and second order mean and variance approximations in these cases. In chapter 5 we include Monte Carlo studies to test how well the moment approximations work in small to moderate sample sizes. We present a small simulation study of nonparametric estimates of variance obtained through the use of the above mentioned variance approximation formulas. The relation of these estimates to the delta method and the bootstrap is noted. Finally, we try one method of extending our theory to quantiles and trimmed means to demonstrate some of the limitations of our results.

Bickel (1967) uses convergence of moments of quantiles and theory on Brownian bridges as his primary tools for showing the convergence of moments of L-estimates. We use similar results on quantiles, but using functional differentiation as our other basic tool allows us to extend Bickel's work in several ways. First, we have fewer restrictions on the distribution function to get convergence of moment results for L-estimates. There is a tradeoff between restrictions on the distribution function and restrictions on the weight function for an L-estimate in formulating theorems on the asymptotics of an L-estimate. Bickel proves results with fewer restrictions on the weight function. Second, we extend his results to higher order moment expansions. Third, our results go beyond his in that we apply them to M-estimates and have the potential to apply them to other robust estimates.

Stigler (1974) has shown that the variances of many L-estimates converge to those of their limiting distributions. His method of proof is to use Hájek projections, which requires L^2-convergence to get convergence in distribution results. The basic assumptions needed are a smooth weight function and either the existence of a variance of the underlying distribution or the deletion of a proportion of the extreme order statistics. Our theorem on L-estimates is an extension to higher moments and higher order expansions of his theorem 5 which has weaker conditions on the tails of the underlying distribution.

Mason (1981) extends Stigler's results on the convergence of variances of L-estimates in the case where the variance of the underlying distribution does not exist. Instead of requiring that a positive proportion of the extreme order statistics have coefficient zero, he requires only that a finite number of the extreme order statistics have coefficient zero. Although we have not done so, it should be possible to extend our results in this fashion.

Eynon (1982) applies some of the theory given here in a study of location and scale invariant M-estimates and P-estimates (P-estimates are analogs of Pitman estimates; see Johns (1979)) using

some of the theory presented here. He has also attempted to automate much of the algebra and calculus on which we spend considerable energy in chapter 4.

Chapter 2

Convergence of moments of quantiles

§2.1 Introduction.

Let $X_1, X_2, \ldots$ be independent, identically distributed (iid) random variables with cumulative distribution function (cdf) F. Our convention will be that F is right continuous; *i.e.*, $F(x) = \mathrm{P}\{X_1 \leq x\}$. Assume $c \in (0,1)$, $F(q) = c$, and the derivative of F at q is $f(q) > 0$. Denote the order statistics of $X_1, X_2, \ldots, X_n$ by $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$. Let $a_n = cn + O(1)$. Assume $Z \sim N(0, c(1-c)/f^2(q))$. Given these assumptions, we will consider conditions on the tails of F which are necessary and sufficient for $\mathrm{E}[g(\sqrt{n}(X_{a_n:n} - q))]$ to converge to $\mathrm{E}[g(Z)]$. We will consider functions g in a class which includes $g(x) = I_{(x_1,\infty)}(x)$, $g(x) = x^r$ and $g(x) = e^{tx}$. The corresponding results are convergence in distribution, convergence of moments, and convergence of the moment generating function, respectively. One pair of necessary and sufficient conditions for g to be in this class is

$$\liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log(|g(x)|)} > 0 \quad \text{and} \quad \liminf_{x\to\infty} \frac{-\log(F(-x))}{\log(|g(-x)|)} > 0. \tag{2.1.1}$$

Another necessary and sufficient condition is that there exists $\delta > 0$ such that $\mathrm{E}[|g(X_1)|^\delta] < \infty$.

A theorem which will eventually connect these two types of tail conditions will be given in section 2.2. This theorem is actually of some interest in itself as it may be used to determine whether or not the expectation of a function of a random variable exists. The results stated above will be proved in section 2.3. In section 2.4 we will discuss other conditions related to the tail conditions in (2.1.1). The moment convergence results will be extended to finite linear combinations of quantiles at the end of section 2.3 and to many robust statistics in chapters 3, 4 and 5. Besides applications to moments and moment generating functions of quantiles from sequences of iid random

variables, applications to sequential occupancy and related problems are suggested by Holst (1981), and Anderson, Sobel and Uppuluri (1982).

Smirnov (1952) and Wretman (1978) have (independently) shown that $\sqrt{n}(X_{a_n:n} - q)$ is asymptotically normal if F is differentiable at q. This is also a simple consequence of a result on Bahadur representation given by Ghosh (1971). Asymptotic theory for the case $f(q) = 0$, the case where left and right hand derivatives of F at q differ, and the case with $a_n = cn + o(\sqrt{n})$ will not be given here. It should be easy to extend the present theory to these cases. Smirnov (1952) gives asymptotic distribution theory for these cases.

Other authors have considered asymptotic behavior of moments of order statistics. The differences in the present treatment are that we consider weaker conditions on F and have introduced tail conditions which are equivalent to other conditions which have been used. Sen (1959) assumes F is continuous everywhere (for convenience) and twice differentiable in some neighborhood of q. The second derivative of F at q is needed to obtain a convergence rate. Bickel (1967) assumes that f is continuous and strictly positive on $\{x : 0 < F(x) < 1\}$ and shows that there exists a $o(n^{-r/2})$ bound for $\mathrm{E}[(\sqrt{n}(X_{a_n:n} - q))^r] - \mathrm{E}[Z^r]$ which is independent of c. He remarks that for c fixed the only local requirement on F needed is that f be continuous in a neighborhood of c—this is more than we require. More discussion on tail conditions and local conditions on F is given in section 2.4.

Bounds for moments of $X_{i:n}$ have been given by various authors. Many such results are summarized in David (1980). In general, one must do more calculation and/or make more assumptions to get these bounds than to get the moment convergence results given here. David also summarizes work done on higher order expansions of moments of order statistics. Of the work presented there, the work of David and Johnson (1954) has the closest relation to our work. Their work is somewhat heuristic in that they do not give tail conditions necessary for their results.

§2.2 Tails and expectations.

A 'nearly' necessary and sufficient condition for the existence of the mean of distribution will be given in this section. Results on the existence of moments and the moment generating function will be considered as examples of extensions of this result. Finally, we extend these results by considering the moments and the moment generating function of an order statistic.

Theorem 2.2.1 provides the main result needed to determine when moments of (functions of) standardized quantiles do not converge. Although it is similar to known results, we have not found the result in the literature.

Theorem 2.2.1. *Suppose X is an arbitrary non-negative random variable with distribution function F. Let*

$$\alpha = \liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log x}.$$

If $\alpha < 1$ then $\mathrm{E}[X] = \infty$. If $\alpha > 1$ then $\mathrm{E}[X] < \infty$. If $\alpha = 1$ either $\mathrm{E}[X] = \infty$ or $\mathrm{E}[X] < \infty$ may hold.

Proof: The last part of the proposition will be proved first. If $F(x) = 1 - 1/x$ for $x \geq 1$ then $\alpha = 1$ and $\mathrm{E}[X] = \infty$. If $F(x) = 1 - x^{-1}e^{-\sqrt{\log x}}$ for $x \geq 1$ then $\alpha = 1$ and $\mathrm{E}[X] = 3$.

Assume $\alpha > 1$. Then there exists $\lambda \in (1, \alpha)$ and x_1 such that if $x \geq x_1$ then $-\log(1-F(x)) > \lambda \log x$. This implies that if $x \geq x_1$ then $1 - F(x) < x^{-\lambda}$ and thus $\mathrm{E}[X] < \infty$.

Now suppose $\alpha < 1$. It will be shown that $\sum_{k=1}^{\infty} \mathrm{P}\{X > k\} = \infty$ which implies $\mathrm{E}[X] = \infty$. If $\alpha < 1$ then there exists $\lambda \in (\alpha, 1)$ and $x_1 < x_2 < \cdots$ such that $1 - F(x_n) > x_n^{-\lambda}$, $n = 1, 2, \cdots$, and $x_n \to \infty$ as $n \to \infty$. Let y_n be the greatest integer less that or equal to x_n, $n = 1, 2, \cdots$, and let $y_0 = 0$. Then

$$\sum_{k=1}^{\infty} \mathrm{P}\{X > k\} = \sum_{n=1}^{\infty} \sum_{k=y_{n-1}+1}^{y_n} \mathrm{P}\{X > k\} \geq \sum_{n=1}^{\infty} \frac{y_n - y_{n-1}}{x_n^{\lambda}} \geq \lim_{n\to\infty} \frac{y_n}{x_n^{\lambda}} = \infty.$$

∎

Definition 2.2.2. *For an arbitrary non-decreasing function g let $g^{-1}(x) = \inf\{y : g(y) \geq x\}$.*

Corollary 2.2.3. *Let X be an arbitrary random variable and denote its distribution function by F. Let g be an arbitrary non-decreasing, non-negative function such that $g(x) \to \infty$ as $x \to F^{-1}(1)$. Let*

$$\alpha = \liminf_{x\to F^{-1}(1)} \frac{-\log(1-F(x))}{\log g(x)}.$$

If $\alpha > 1$ then $\mathrm{E}[g(X)] < \infty$, if $\alpha < 1$ then $\mathrm{E}[g(X)] = \infty$, and if $\alpha = 1$ either $\mathrm{E}[g(X)] = \infty$ or $\mathrm{E}[g(X)] < \infty$ may hold.

Proof: Let $1 < y < \infty$ and let $x = g^{-1}(y)+$. Then $g(x) \geq y$ and

$$\frac{-\log(1-F(x))}{\log g(x)} \leq \frac{-\log(1-F(g^{-1}(y)))}{\log y}.$$

Since as $y \to \infty$, x chosen in this fashion goes to $F^{-1}(1)$ it follows that

$$\liminf_{x\to F^{-1}(1)} \frac{-\log(1-F(x))}{\log g(x)} \leq \liminf_{y\to\infty} \frac{-\log(1-F(g^{-1}(y)))}{\log y}.$$

If we fix $x \in (g^{-1}(1+), F^{-1}(1))$ and let $y = g(x)$ then $g^{-1}(y) \leq x$ and

$$\frac{-\log(1-F(x))}{\log g(x)} \geq \frac{-\log(1-F(g^{-1}(y)))}{\log y}.$$

As $x \to F^{-1}(1)$, y chosen in this fashion goes to ∞ and thus

$$\liminf_{x \to F^{-1}(1)} \frac{-\log(1-F(x))}{\log g(x)} \geq \liminf_{y \to \infty} \frac{-\log(1-F(g^{-1}(y)))}{\log y}.$$

We have now shown

$$\liminf_{x \to F^{-1}(1)} \frac{-\log(1-F(x))}{\log g(x)} = \liminf_{y \to \infty} \frac{-\log(1-F(g^{-1}(y)))}{\log y}.$$

Since

$$\begin{aligned}\liminf_{x \to \infty} \frac{-\log(1-F(g^{-1}(x)))}{\log x} &= \liminf_{x \to \infty} \frac{-\log(1-F(g^{-1}(x+)))}{\log x} \\ &= \liminf_{x \to \infty} \frac{-\log \mathrm{P}\{g(X) > x\}}{\log x}\end{aligned}$$

the contention follows from theorem 2.2.1. ∎

Corollary 2.2.4. *For an arbitrary random variable X with distribution function F let*

$$\alpha_+ = \liminf_{x \to \infty} \frac{-\log(1-F(x))}{\log x}, \quad \alpha_- = \liminf_{x \to \infty} \frac{-\log F(-x)}{\log x},$$

and $\alpha = \min(\alpha_+, \alpha_-)$. If $0 \leq r < \alpha$ then $\mathrm{E}[|X|^r] < \infty$. If $r > \alpha$ then $\mathrm{E}[|X|^r] = \infty$. If $r = \alpha > 0$ then either $\mathrm{E}[|X|^r] = \infty$ or $\mathrm{E}[|X|^r] < \infty$ may hold.

Corollary 2.2.5. *For an arbitrary random variable X with distribution function $F(x)$ let*

$$\alpha_2 = \liminf_{x \to \infty} \frac{-\log(1-F(x))}{x} \quad \textit{and} \quad \alpha_1 = -\liminf_{x \to \infty} \frac{-\log F(-x)}{x}.$$

If $t \in (\alpha_1, \alpha_2)$ then $\mathrm{E}[e^{tX}] < \infty$. If $t < \alpha_1$ or $t > \alpha_2$ then $\mathrm{E}[e^{tX}] = \infty$.

Note that x is in the denominator of the functions defining α_1 and α_2 whereas $\log x$ is in the denominator when defining α_+ and α_-. The values α_+ and α_- are equally useful when considering the existence of moments of order statistics.

Theorem 2.2.6. *Define α_+ and α_- as in corollary 2.2.4. If $i > r/\alpha_-$ and $n - i + 1 > r/\alpha_+$ then $\mathrm{E}[|X_{i:n}|^r] < \infty$. If $i < r/\alpha_-$ or $n - i + 1 < r/\alpha_+$ then $\mathrm{E}[|X_{i:n}|^r] = \infty$.*

Proof: First assume $n - i + 1 < r/\alpha_+$. Since

$$\mathrm{P}\{|X_{i:n}|^r > x\} \geq \mathrm{P}\{X_{i:n} > x^{1/r}\} \geq (1 - F(x^{1/r}))^{n-i+1},$$

and since

$$\liminf_{x\to\infty} \frac{-\log((1-F(x^{1/r}))^{n-i+1})}{\log x} = \frac{n-i+1}{r} \liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log x} < 1 \tag{2.2.1}$$

it follows from theorem 2.2.1 that $\mathrm{E}[|X^r_{i:n}|] = \infty$. For $i < r/\alpha_-$ the proof is analogous.

Now assume $r/\alpha_- < i$ and $r/\alpha_+ < n-i+1$. We have

$$\begin{aligned}&\liminf_{x\to\infty} \frac{-\log \mathrm{P}\{|X_{i:n}|^r > x\}}{\log x}\\ &\geq \min\left(\liminf_{x\to\infty} \frac{-\log \mathrm{P}\{X_{i:n} < -x^{1/r}\}}{\log x}, \liminf_{x\to\infty} \frac{-\log \mathrm{P}\{X_{i:n} > x^{1/r}\}}{\log x}\right).\end{aligned} \tag{2.2.2}$$

Since

$$\mathrm{P}\{X_{i:n} > x^{1/r}\} \leq \binom{n}{i}(1-F(x^{1/r}))^{n-i+1},$$

the inequality in (2.2.1) can be switched in this case to $>$ and we have

$$\liminf_{x\to\infty} \frac{-\log \mathrm{P}\{X_{i:n} > x^{1/r}\}}{\log x} > 1$$

Similarly,

$$\liminf_{x\to\infty} \frac{-\log \mathrm{P}\{X_{i:n} < -x^{1/r}\}}{\log x} > 1$$

and from (2.2.2) and theorem 2.2.1 the contention follows. ∎

§2.3 Convergence of moments of standardized quantiles.

We are now prepared to address the questions of interest. The proof of the most general moment convergence result for quantiles is analytic and somewhat tedious in nature. The proof of convergence does not use uniform integrability. A proof using uniform integrability would require analysis similar to that given below. The proof of the necessity of tail conditions such as (2.1.1) for the existence of moments of standardized quantiles will now become trivial in many cases. Following is the most general result concerning 'necessity' that will be given. It is a variation of theorem 2.2.6. In applications we will take $n > 1$ and β (defined in the theorem) to be $\sqrt{n}$.

Theorem 2.3.1. *Suppose g is a non-decreasing, non-negative function, and $X_1, X_2, \ldots$ are independent identically distributed random variables with distribution function F where $F^{-1}(1) = \infty$. Suppose further that*

$$\liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log g(x)} = 0.$$

Then for any $\beta > 1$, real η, and positive integers n, i, $1 \le i \le n$, it follows that

$$\mathrm{E}[g(\beta(X_{i:n} - \eta))] = \infty.$$

Proof: Since for large x

$$\begin{aligned} \mathrm{P}\{g(\beta(X_{i:n} - \eta)) > x\} &\ge \mathrm{P}\{g(X_{i:n}) > x\} \\ &\ge (1 - F(g^{-1}(x+)))^{n-i+1} \end{aligned}$$

it follows that

$$\liminf_{x\to\infty} \frac{-\log \mathrm{P}\{g(\beta(X_{i:n} - \eta)) > x\}}{\log x} \le (n-i+1) \liminf_{x\to\infty} \frac{-\log(1 - F(g^{-1}(x+)))}{\log x}.$$

From the proof of corollary 2.2.3 we can see that

$$\liminf_{x\to\infty} \frac{-\log(1 - F(g^{-1}(x+)))}{\log x} = \liminf_{x\to\infty} \frac{-\log(1 - F(x))}{\log g(x)}$$

and the proof is completed by the application of theorem 2.2.1. ∎

We will now address the problem of convergence. It will be useful to label the following assumptions:

i. g is a finite, continuous, non-decreasing, non-negative function defined for all real numbers.

ii. g is bounded, or there exist $\beta, x_0 > 0$ such that if $t > 1$, $x > x_0$ then $\log g(tx) < t\beta \log g(x)$.

iii. $\liminf_{x\to\infty} \frac{-\log(1 - F(x))}{\log g(x)} > 0$.

iv. $c \in (0,1)$, $a_n = cn + O(1)$, $c_n = a_n/n = c + O(1/n)$.

v. $F(q) = c$, $\frac{d}{dx}F(x)\,|_{x=q} = f(q) > 0$.

We have chosen assumptions *i* and *ii* to make the proofs of our general results simple. The functions g in which we are interested are $I_{[y,\infty)}(x)$ (this does not satisfy condition *i*, but this problem can be overcome by smoothing), $x^r I_{[0,\infty)}(x)$, $r = 1, 2, \ldots$, and e^{tx}. Clearly e^{tx} for $t \ge 0$ and $I_{[y,\infty)}(x)$ satisfy *ii*. To show that $x^r I_{[0,\infty)}(x)$ satisfies *ii* we let $x > x_0 > e$. For $t \ge 1$ we define $h_1(t) = r\log(tx)$. Then $h_1(1) = r\log x$ and for $t \ge 1$, $h_1'(t) = r/t \le r$. Now define $h_2(t) = rt\log x$. Then $h_2(1) = r\log x$ and for $t \ge 1$, $h_2'(t) = r\log x > r$. These facts imply that if $t > 1$ then $h_1(t) < h_2(t)$. It now follows that if $x_0 > e$ then *ii* holds for $x^r I_{[0,\infty)}(x)$. More discussion on assumption *ii* will be given in section 2.4.

To avoid complicated formulas for cases with F not continuous we use the representation

$$\mathrm{E}[g(\sqrt{n}(X_{a_n:n}-q))]=\int_0^1 g(\sqrt{n}(F^{-1}(u)-q))n\binom{n-1}{a_n-1}u^{a_n-1}(1-u)^{n-a_n}du. \tag{2.3.1}$$

Using Stirling's formula $n!=\sqrt{2\pi}n^{n+1/2}e^{-n}e^{\theta_n/(12n)}$ where $0<\theta_n<1$ (see, *e.g.*, Rényi (1970), p. 149 ff.) and letting $c_n=c+O(1/n)$ as in *iv* it is a straightforward calculation to show that as $n\to\infty$

$$\binom{n-1}{a_n-1}=(\sqrt{2\pi n})^{-1}c_n^{-nc_n+1/2}(1-c_n)^{-n(1-c_n)-1/2}(1+O(1/n)). \tag{2.3.2}$$

Thus uniformly for $u\in(0,1)$

$$n\binom{n-1}{a_n-1}u^{a_n-1}(1-u)^{n-a_n}\sim\sqrt{\frac{nc_n}{2\pi(1-c_n)}}\frac{1}{u}\left(\frac{u}{c_n}\right)^{nc_n}\left(\frac{1-u}{1-c_n}\right)^{n(1-c_n)} \tag{2.3.3}$$

Let

$$p(u,v)=v(\log v-\log u)+(1-v)(\log(1-v)-\log(1-u)). \tag{2.3.4}$$

In the following we let $\log 0=-\infty$ and $e^{-\infty}=0$. From this and (2.3.1)—(2.3.4) it follows if g is non-negative and $c_n=c+O(1/n)$ then as $n\to\infty$

$$\mathrm{E}[g(\sqrt{n}(X_{a_n:n}-q))]\sim\int_0^1\sqrt{\frac{nc_n}{2\pi(1-c_n)}}\frac{1}{u}\exp\left(\log g(\sqrt{n}(F^{-1}(u)-q))-np(u,c_n)\right)du. \tag{2.3.5}$$

The proof of the following lemma contains the key ideas needed to show convergence of expectations of functions of standardized quantiles.

Lemma 2.3.2. *Under assumptions i—iv if $\epsilon,b>0$ and $\gamma<1/4$ then*

$$0=\lim_{n\to\infty}\int_{c+\epsilon n^{-\gamma}}^1 n^b\frac{1}{u}\exp\left(\log g(\sqrt{n}(F^{-1}(u)-q))-np(u,c_n)\right)du.$$

Proof: We choose an arbitrary $\epsilon>0$. Without loss of generality we assume $\gamma\in(0,1/4)$ and $q\le F^{-1}(c)$.

First we will show that the integrand goes pointwise to zero. We will consider a Taylor's series approximation of p in the neighborhood of (c,c). The first and second partial derivatives of p are

$$\frac{\partial p}{\partial v}=\log v-\log u-\log(1-v)+\log(1-u), \tag{2.3.6}$$

$$\frac{\partial p}{\partial u}=\frac{-v}{u}+\frac{1-v}{1-u}, \tag{2.3.7}$$

$$\frac{\partial^2 p}{\partial v^2} = \frac{1}{v} + \frac{1}{1-v} > 0,$$

$$\frac{\partial^2 p}{\partial u^2} = \frac{v}{u^2} + \frac{1-v}{(1-u)^2} > 0, \quad \text{and}$$

$$\frac{\partial^2 p}{\partial u \partial v} = -\frac{1}{u} - \frac{1}{1-u}.$$

Using the first order Taylor's series expansion with remainder, it follows from the above that for some $\theta_1(u,v)$ in the closed interval between c and u and $\theta_2(u,v)$ in the closed interval between c and v

$$p(u,v) = \frac{(u-c)^2}{2}\left(\frac{\theta_2(u,v)}{\theta_1^2(u,v)} + \frac{1-\theta_2(u,v)}{(1-\theta_1(u,v))^2}\right) + \frac{(v-c)^2}{2}\left(\frac{1}{\theta_2(u,v)} + \frac{1}{1-\theta_2(u,v)}\right)$$
$$- (v-c)(u-c)\left(\frac{1}{\theta_1(u,v)} + \frac{1}{1-\theta_1(u,v)}\right).$$

Since $c_n = c + O(1/n)$ and $p(u,v)$ (from (2.3.7)) is increasing in u for $u > v$ we have for some N_1, all $n \geq N_1$, and $u \in [c+\epsilon n^{-\gamma}, 1)$

$$p(u, c_n) \geq c\epsilon^2 n^{-2\gamma}/4.$$

Thus for $n \geq N_1$ the integrand is bounded by

$$c^{-1}\exp\left(b\log n + \log g(\sqrt{n}(F^{-1}(u) - q)) - n^{1-2\gamma}c\epsilon^2/4\right). \tag{2.3.8}$$

For g bounded the result follows immediately from the dominated convergence theorem since (2.3.8) goes pointwise to zero and since if $n \geq N_1$ this may be bounded by some constant for all u. Similarly, the result now follows for the case with $q = F^{-1}(c) = F^{-1}(1)$.

Now assume that g is not bounded and $F^{-1}(1) = \infty$. Let $c_0 = \sup_{n \geq N_1} c_n$. We suppose N_1 is sufficiently large so that $c_0 < 1$. It follows from (2.3.6) that if $n \geq N_1$ and $u \geq c_0$ then

$$p(u, c_n) \geq p(u, c_0). \tag{2.3.9}$$

From the assumption that $\liminf_{x\to\infty}(-\log(1-F(x))/\log g(x)) > 0$ (assumption *iii*) we see that there exist $u_1 \in (c_0, 1)$, $\eta_1 > 0$ such that if $u \in (u_1, 1)$ then

$$-\log(1-u) > \eta_1 \log g(F^{-1}(u)).$$

From (2.3.4) we see that there exists $\eta_2 > 0$ such that for $u \in (u_1, 1)$

$$p(u, c_0) > -\eta_2 \log(1-u).$$

Thus for $u \in (u_1, 1)$ and $\eta = \eta_1\eta_2 > 0$

$$p(u, c_0) > \eta \log(g(F^{-1}(u))). \tag{2.3.10}$$

We now assume in addition to the above that u_1 is so large that $F^{-1}(u_1) - q > x_0$ where x_0 is as in *ii*. We let $k > 1$ (this is needed below because q may be less that zero) be such that if $u \in (u_1, 1)$ and $n \geq N_1$ then

$$\log g(\sqrt{n}(F^{-1}(u) - q)) < k\sqrt{n}\beta \log g(F^{-1}(u)). \tag{2.3.11}$$

From (2.3.9)—(2.3.11) we may now bound the integrand of the lemma for $u \in (u_1, 1)$, $n \geq N_1$ by

$$c^{-1} \exp\left(b \log n + (k\sqrt{n}\beta - n\eta) \log g(F^{-1}(u))\right).$$

Given this bound we may now apply the dominated convergence theorem to the integral on the interval $(u_1, 1)$.

If $F^{-1}(1) < \infty$ and $q < F^{-1}(1)$ let $u_1 = 1$. Otherwise let u_1 be as in the previous paragraph. Let x_0 be as in *ii*. Let $N_2 \geq N_1$ be such that $\sqrt{N_2}(F^{-1}(u_1) - q) > x_0$. Assumption *ii* then implies that for $n \geq N_2$ and $c + \epsilon n^{-\gamma} \leq u \leq u_1$

$$\begin{aligned}\log g(\sqrt{n}(F^{-1}(u) - q)) &\leq \log g(\sqrt{n}(F^{-1}(u_1) - q)) \\ &\leq \beta\sqrt{n/N_2} \log g(\sqrt{N_2}(F^{-1}(u_1) - q)),\end{aligned}$$

which with (2.3.8) implies that the integrand is bounded on $[c + \epsilon n^{-\gamma}, u_1]$ for $n \geq N_2$ by

$$c^{-1} \exp\left(b \log n + \beta\sqrt{n/N_2} \log g(\sqrt{N_2}(F^{-1}(u_1) - q)) - n^{1-2\gamma}c\epsilon^2/4\right).$$

It follows that the integral on $(c + \epsilon n^{-\gamma}, u_1]$ goes to zero as $n \to \infty$ by the dominated convergence theorem, and the proof is complete. ∎

One more result is needed to show convergence of moments. We could choose to show uniform integrability of $g(\sqrt{n}(X_{a_n:n} - q))$ and then use Wretman's (1978) result for convergence in distribution. However, it is almost as easy to show convergence of moments directly using lemma 2.3.3 given below. The constant 2/9 of the lemma is arbitrary. The proof holds when this value is replaced by any number between 1/5 and 1/2. A constant less than 1/4 is needed to use the lemma in conjunction with lemma 2.3.2.

Lemma 2.3.3. *Assume conditions i,ii, and iv hold, $k > 0$, and $Y \sim N(0, k^2c(1-c))$. Then for any $\epsilon > 0$ as $n \to \infty$*

$$\xi_n(\epsilon) = \int_{c-\epsilon/n^{2/9}}^{c+\epsilon/n^{2/9}} g(k\sqrt{n}(u-c))n\binom{n-1}{a_n-1}u^{a_n-1}(1-u)^{n-a_n}du \to \mathrm{E}[g(Y)].$$

Proof: Assume $k=1$. By applying equation (2.3.3) and letting $z=\sqrt{n}(u-c_n)$ we see that

$$\xi_n(\epsilon) \sim \int_{-\epsilon n^{5/18}+\sqrt{n}(c-c_n)}^{\epsilon n^{5/18}+\sqrt{n}(c-c_n)} \frac{g(z+\sqrt{n}(c_n-c))}{\sqrt{2\pi c(1-c)}}\left(1+\frac{z}{\sqrt{n}c_n}\right)^{a_n-1}\left(1-\frac{z}{\sqrt{n}(1-c_n)}\right)^{n-a_n} dz.$$

We will use the Taylor's series expansion

$$\log(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\frac{x^5}{5(1+\theta(x))^5}$$

which is valid for $-1<x<1$ and some $\theta(x)$ between 0 and x. Let

$$w_n(z)=\frac{(1-2c)z^3}{3\sqrt{n}c^2(1-c)^2}-\frac{(1-3c+3c^2)z^4}{4nc^3(1-c)^3}.$$

It is a straightforward calculation to show that as $n\to\infty$

$$\max_{|z|<\epsilon n^{5/18}+\sqrt{n}|c-c_n|}\left|\log\left(\left(1+\frac{z}{\sqrt{n}c_n}\right)^{a_n-1}\left(1-\frac{z}{\sqrt{n}(1-c_n)}\right)^{n-a_n}\right)+\frac{z^2}{2c(1-c)}-w_n(z)\right|\to 0.$$

Thus

$$\xi_n(\epsilon)\sim\int_{\epsilon n^{5/18}+\sqrt{n}(c-c_n)}^{\epsilon n^{5/18}+\sqrt{n}(c-c_n)}\frac{g(z+\sqrt{n}(c_n-c))}{\sqrt{2\pi c(1-c)}}\exp\left(-z^2/(2c(1-c))+w_n(z)\right)dz.$$

Since $c_n=c+O(1/n)$ assumption *i* implies $g(z+\sqrt{n}(c_n-c))\to g(z)$ for each fixed z. Assumption *ii* implies that g grows at most exponentially. Since $z/\sqrt{n}=o(1)$ uniformly for all z in the range of integration it follows that $w_n(z)=o(z^2)$ uniformly in this range. These facts imply that the integrand above can be bounded by $k_1\exp(-k_2z^2)$ for some k_1, k_2 and n large. Thus by the dominated convergence theorem $\xi_n(\epsilon)\to \mathrm{E}[g(Y)]$.

For $k\neq 1$ we let $g_1(x)=g(kx)$ and apply the result for $k=1$ to g_1 since *i*, *ii*, and *iv* are still satisfied. ∎

We are now prepared to prove the most general convergence result which we shall present.

Theorem 2.3.4. *If conditions i—v are satisfied and* $Z\sim N(0,c(1-c)/f^2(q))$, *then as* $n\to\infty$

$$\mathrm{E}[g(\sqrt{n}(X_{a_n:n}-q))]\to \mathrm{E}[g(Z)].$$

Proof: From (2.3.1), (2.3.3), and lemma 2.3.2 and its obvious analog it follows that for any $\epsilon>0$,

$$\mathrm{E}[g(\sqrt{n}(X_{a_n:n}-q))]\sim\int_{c-\epsilon/n^{2/9}}^{c+\epsilon/n^{2/9}} g(\sqrt{n}(F^{-1}(u)-q))n\binom{n-1}{a_n-1}u^{a_n-1}(1-u)^{n-a_n}du.$$

For any $\delta>0$ there exists $\epsilon>0$ such that for $u\in(c-\epsilon,c+\epsilon)$

$$(1-\delta)\frac{u-c}{f(q)}<F^{-1}(u)-q<(1+\delta)\frac{u-c}{f(q)}. \qquad (2.3.12)$$

The result now follows from lemma 2.3.3 by substituting the two bounds into the integrand above and noting that $\delta>0$ was arbitrary. ∎

We will now consider applications of theorems 2.3.1 and 2.3.4. For propositions 2.3.5—2.3.9 assume that *iv* and *v* hold, and that $Z \sim N(0, c(1-c)/f^2(q))$.

Proposition 2.3.5. *As $n \to \infty$, $\sqrt{n}(X_{a_n:n} - q)$ converges in distribution to Z.*

Proof: We will show that for x_1 arbitrary $\mathrm{P}\{\sqrt{n}(X_{a_n:n} - q) \geq x_1\} \to \mathrm{P}\{Z \geq x_1\}$ as $n \to \infty$. For $\epsilon > 0$ we let

$$g_\epsilon(x) = I_{[x_1,\infty)}(x) + (1 + (x - x_1)/\epsilon) I_{[x_1-\epsilon,x_1]}(x).$$

and apply theorem 2.3.4 to g_ϵ. By letting $\epsilon \to 0$ we obtain

$$\limsup_{n\to\infty} \mathrm{P}\{\sqrt{n}(X_{a_n:n} - q) \geq x_1\} \leq \mathrm{P}\{Z \geq x_1\}.$$

In a similar fashion we can show

$$\liminf_{n\to\infty} \mathrm{P}\{\sqrt{n}(X_{a_n:n} - q) \geq x_1\} \geq \mathrm{P}\{Z \geq x_1\},$$

and the proof is complete. ∎

Proposition 2.3.6. *Define α_+ and α_- as in corollary 2.2.4. If r is a positive integer then $\mathrm{E}[(\sqrt{n}(X_{a_n:n} - q))^r] \to \mathrm{E}[Z^r]$ as $n \to \infty$ if and only if $\alpha_+ > 0$ and $\alpha_- > 0$.*

Proof: The necessity of $\alpha_+ > 0$ and $\alpha_- > 0$ follows from theorem 2.2.6 or theorem 2.3.1.

Now assume $\alpha_+ > 0$ and $\alpha_- > 0$. Letting $g(x) = x^r I_{\{x>0\}}(x)$ it follows from theorem 2.3.4 that $\mathrm{E}[g(\sqrt{n}(X_{a_n:n} - q))] \to \mathrm{E}[g(Z)]$. Letting $Y_i = -X_i$, $b_n = n + 1 - a_n$ it follows that $\mathrm{E}[g(\sqrt{n}(Y_{b_n:n} - q))] \to \mathrm{E}[g(Z)]$. The result now follows since

$$\mathrm{E}[(\sqrt{n}(X_{a_n:n} - q))^r] = \mathrm{E}[g(\sqrt{n}(X_{a_n:n} - q))] + (-1)^r E[g(\sqrt{n}(Y_{b_n:n} - q))].$$

∎

Proposition 2.3.7 below follows from proposition 2.3.6 and corollary 2.2.4. Proposition 2.3.8 follows from theorems 2.3.1 and 2.3.4. Proposition 2.3.9 follows from proposition 2.3.8 and corollary 2.2.5.

Proposition 2.3.7. *If $r > 0$ then $\mathrm{E}[(\sqrt{n}(X_{a_n:n} - q))^r] \to \mathrm{E}[Z^r]$ as $n \to \infty$ if and only if there exists $\delta > 0$ such that $\mathrm{E}[|X_1|^\delta] < \infty$.*

Proposition 2.3.8. *Define α_1 and α_2 as in corollary 2.2.5. Then* $\mathrm{E}[\exp(t\sqrt{n}(X_{a_n:n} - q))] \to \mathrm{E}[e^{tZ}]$ *for all $t > 0$ if and only if $\alpha_2 > 0$. Similarly* $\mathrm{E}[\exp(t\sqrt{n}(X_{a_n:n} - q))] \to \mathrm{E}[e^{tZ}]$ *for all $t < 0$ if and only if $\alpha_1 < 0$.*

Proposition 2.3.9. *For all $t > 0$* $\mathrm{E}[\exp(t\sqrt{n}(X_{a_n:n} - q))] \to \mathrm{E}[e^{tZ}]$ *if and only if there exists $\epsilon > 0$ such that* $\mathrm{E}[e^{\epsilon X_1}] < \infty$. *Similarly* $\mathrm{E}[\exp(t\sqrt{n}(X_{a_n:n} - q))] \to \mathrm{E}[e^{tZ}]$ *for all $t < 0$ if and only if there exists $\epsilon > 0$ such that* $\mathrm{E}[e^{-\epsilon X_1}] < \infty$.

In the next section we will note some relations that sometimes simplify applications of the above propositions. These arguments and propositions 2.3.6 and 2.3.7 imply that moments of standardized quantiles from all commonly considered distributions converge. Propositions 2.3.8 and 2.3.9 imply that the moment generating functions of standardized quantiles from distributions with exponential tails converge to the moment generating function of a normal distribution. They also imply that moment generating functions for standardized quantiles for distributions such as the Cauchy, Pareto and slash distributions do not converge; in fact from theorem 2.3.1 it follows that they do not exist for any n.

Before closing this section we will extend our result on convergence of moments of quantiles to finite linear combinations of quantiles. We will need the following lemma which is a trivial extension of theorem 4.5.2 of Chung (1974).

Lemma 2.3.10. *If Y_n, $n = 1, 2, \ldots$ converges in distribution to X, and for some $p > 0$,* $\limsup_{n\to\infty} \mathrm{E}[|Y_n|^p] = M < \infty$ *then for each $r < p$*

$$\lim_{n\to\infty} E[|Y_n|^r] = E[|Y|^r] < \infty.$$

If r is a positive integer, then we may replace $|Y_n|^r$ and $|Y^r|$ above by Y_n^r and Y^r, respectively.

Proposition 2.3.11. *Let α_+ and α_- be as in corollary 2.2.4. Suppose F is a cdf, $0 < c_1 < \ldots < c_k < 1$, $c_i = F(q_i)$, and $(d/dx)F(x)\,|_{x=q_i} = f(q_i) > 0$, $1 \le i \le k$. Suppose further that $a_{in} = nc_i + O(1)$, $1 \le i \le k$. Finally, let Z_i, $1 \le i \le k$ have a multivariate normal distribution with $\mathrm{E}[Z_i] = 0$, $\mathrm{E}[Z_iZ_j] = c_i(1 - c_j)/(f(q_i)f(q_j))$, $1 \le i \le j \le k$. Then for any finite constants b_i, $1 \le i \le k$, $r = 1, 2, \ldots$*

$$\lim_{n\to\infty} \mathrm{E}\left[\left(\sum_{i=1}^{k} b_i(X_{a_{in}:n} - q_i)\right)^r\right] = \mathrm{E}\left[\left(\sum_{i=1}^{k} b_i Z_i\right)^r\right]$$

if and only if $\alpha_+ > 0$ and $\alpha_- > 0$.

Proof: If $\alpha_+ = 0$ or $\alpha_- = 0$ it follows from theorem 2.2.6 that $\mathrm{E}[|\, X_{i:n}\,|^r] = \infty$ for $1 \le i \le n < \infty$. Thus the necessity of $\alpha_+ > 0$ and $\alpha_- > 0$ is established.

Assume $\alpha_+ > 0$ and $\alpha_- > 0$. First we wish to show that the vector $\sqrt{n}(X_{a_{1n}:n} - q_1, \ldots, X_{a_{kn}:n} - q_k)$ converges in distribution to $(Z_1, \ldots, Z_k)$. The proof of this is the same as that of David (1980), pp. 255-257, for a result with F continuous everywhere. The basic tool used in the proof is the result of Ghosh (1971) on Bahadur representation. From the proof of proposition 2.3.6 we know that for $1 \le i \le k$

$$\limsup_{n\to\infty} E[\sqrt{n}|(X_{a_{in}:n} - q_i)|^r] = E[|Z_i|^r]. \tag{2.3.13}$$

Applying the Minkowski inequality repeatedly we have

$$\begin{aligned}\limsup_{n\to\infty}\left(\mathrm{E}\left[\left|\sum_{i=1}^{k} b_i\sqrt{n}(X_{a_{in}:n} - q_i)\right|^r\right]\right)^{1/r} &\le \limsup_{n\to\infty}\sum_{i=1}^{k}|b_i|(\mathrm{E}[|\sqrt{n}(X_{a_{in}:n} - q_i)|^r])^{1/r} \\ &\le \sum_{i=1}^{k}|b_i|\limsup_{n\to\infty}(\mathrm{E}[\sqrt{n}|(X_{a_{in}:n} - q_i)|^r])^{1/r}.\end{aligned}$$

The proposition now follows from (2.3.13) and lemma 2.3.10. ∎

§2.4 Remarks

In this section we will discuss assumptions *ii* and *iii* of the previous section and their implications. We will also note the local conditions on F which have been used by other authors to show convergence of moments of quantiles.

We define

$$\alpha = \liminf_{x\to\infty}\frac{-\log(1-F(x))}{\log x} \quad \text{and} \quad \alpha' = \liminf_{x\to\infty}\frac{-\log(1-F(x))}{x}. \tag{2.4.1}$$

If $\alpha > 0$ then for any $\lambda \in (0, \alpha)$ there exists x_λ such that if $x > x_\lambda$ then $1 - F(x) < x^{-\lambda}$. Similarly, if $\alpha' > 0$ then for any $\lambda \in (0, \alpha')$ there exists x'_λ such that if $x > x'_\lambda$ then $1 - F(x) < e^{-\lambda x}$. Bickel (1967) has used the existence of some $\epsilon > 0$ such that $\lim_{x\to\infty} x^\epsilon(1 - F(x) + F(-x)) = 0$ to give moment results similar to (but less precise than) theorem 2.2.6 and to get convergence of moment results. Blom (1959), p. 44, has used a bound proportional to $u^a(1-u)^b$ (with $a, b < 0$) for $|F^{-1}(u)|$ as a condition to obtain a result similar to theorem 2.2.6 for first and second moments (this result could be extended to higher moments easily). The results of section 2.2 showing when moments and moment generating functions do and do not exist demonstrate that defining the precise values of α and α' has some utility.

Another concept related to α is regular variation. See, for example, Feller (1971) or de Haan (1970) for definitions and some elementary properties. The statement that $1 - F(x) < kx^{-\alpha}$ for

some $k, \alpha > 0$ and large x implies that $1 - F(x)$ is bounded by a function having regular variation. If $F(x) = 1 - x^{-\alpha}$ standard results in asymptotic extreme value theory imply that if $0 < \alpha < \infty$ then the distribution function of $X_{n:n}/n^{1/\alpha}$ converges vaguely to $1 - e^{-x^{\alpha}}$. Pickands (1968) studied asymptotic behavior of moments of sample extremes. His basic result was that if $b_n X_{n:n}$ has a limiting distribution which has an r^{th} moment then $\mathrm{E}[(b_n X_{n:n})^r]$ converges to that value. This implies that if $0 < \alpha < \infty$ where α is as in (2.4.1) then $\mathrm{E}[X^r_{n:n}] = O(n^{r/\alpha})$.

We may define α (and α') in terms of the quantile function $F^{-1}(u)$; that is

$$\alpha = \liminf_{u \to 1} \frac{-\log(1-u)}{\log F^{-1}(u)}.$$

The value $1/\alpha$ is what Parzen (1979) refers to as the tail exponent of $f(F^{-1}(u))$; *i.e.*, if $f(F^{-1}(u))$ is regularly varying as $u \to 1$, then the exponent of regular variation is $1/\alpha$.

The function $\Lambda(x) \equiv -\log(1 - F(x))$ is often referred to as the cumulative hazard function. Its derivative $\lambda(x) = f(x)/(1 - F(x))$ is the hazard rate. Suppose for large x that $\lambda(x)$ is bounded below by βx^{δ} where $\beta > 0$. If $\delta \geq -1$ then $\alpha > 0$, and if $\delta \geq 0$ then $\alpha' > 0$. On the other hand, suppose $\lambda(x)$ is bounded above for large x by βx^{δ} where $\beta > 0$. If $\delta < -1$ then $\alpha = 0$, and if $\delta < 0$ then $\alpha' = 0$.

Other authors have used stronger local conditions than assumption v to obtain convergence of moments of quantiles. Sen (1959) used the existence of f' in a neighborhood of q to obtain a higher order approximation (than (2.3.12)) of F in that neighborhood. Bickel (1967) used the representation $F^{-1}(u) - q = (u - c)/f(x(u))$ where $x(u)$ is between q and $F^{-1}(u)$ and thus required f continuous in a neighborhood of q.

When g is unbounded condition *ii* may be written as

ii$'$. There exist $\beta, x_0 > 0$ such that if $t > 1, x > x_0$, then $\log g(tx) < t\beta \log g(x)$.

We have chosen this condition to simplify the proof of lemma 2.3.2. This inequality implies that if $x > x_0$ and $t > 1$ then $g(tx) < g(x)^{t\beta}$ which implies that $g(x)$ grows at most exponentially for large x. The crucial implication of *ii*$'$ (which is not implied by the fact that g is bounded by an exponential function) is that

$$(1/2)\log n + \log g(\sqrt{n}(x-q)) - n \log g(x) \to -\infty$$

uniformly for $x > x_1$ for some $x_1 > q$ as $n \to \infty$; *i.e.*, this implies that $\log g(\sqrt{n}(x-q)) < k\sqrt{n}\log g(x)$ for $x > x_1$, n large and some k, which is exactly what we need at the end of the proof of lemma 2.3.2.

Chapter 3

Expansions for moments of robust statistics

§3.1 Introduction.

A theorem showing asymptotic expansions for moments for a class of robust statistics is given in this chapter. Included in this class are many statistics which may be written as a functional of the empirical distribution function; *i.e.*, they may be written as $T(F_n)$ where F_n is an empirical distribution function based on n independent, identically distributed (iid) random variables with underlying cumulative distribution function (cdf) F. The basic result is that the tail condition on F given in chapter 2 which implies convergence of moments of standardized quantiles is found to be sufficient to imply convergence of moments of $\sqrt{n}(T(F_n) - T(F))$ to moments of expansions of $T(F_n)$ about F obtained using a version of Fréchet differentiation. The result gives higher order approximations of moments if the defining functional has higher order Fréchet derivatives.

We begin with two short sections establishing some notation and definitions. Then we present our basic theorem and its proof. All of the theory of Fréchet differentiation needed is presented in sections 4 and 5. In section 5 we apply our basic theorem to give general formulas for first and second order approximations to the mean and mean squared error. Applications of these results to L- and M-estimates are given in the next chapter.

§3.2 Notation.

Much of the notation needed for the remainder of the thesis is given in this and the following section. In this section we present some preliminary notation which will be needed in our discussion of functional statistics and Fréchet differentiation. The definition of Fréchet differentiation and corresponding notation will be given in the next section.

$\Re$ —the space of real numbers.

cdf —cumulative *probability* distribution function.

D —the space of finite linear combinations of one dimensional cdf's.

G (or G_i) —an arbitrary element of D; often a cdf or a difference of cdf's.

F —an arbitrary element of D; usually a cdf.

$X_1, X_2, \ldots$ —an iid sequence with cdf F.

δ_x —the cdf with mass one at x.

$F_n \equiv (1/n)\sum_{i=1}^{n} \delta_{X_i}$ —the empirical cdf after n iid observations from F.

$Z_i \equiv \delta_{X_i} - F$; a D-valued random variable.

$\|\cdot\|$ —an arbitrary norm on D.

$\|\cdot\|_\infty$ —the sup norm on D; *i.e.*, for $G \in D$, $\| G \|_\infty \equiv \sup_{-\infty<x<\infty} | G(x) |$.

$D_n \equiv \| F_n - F \|_\infty$, the Kolmogorov-Smirnov statistic after n observations.

§3.3 Functional differentiation and von Mises expansions.

Definitions pertaining to functional differentiation which will be needed are collected in this section. We consider only Fréchet differentiation as we need bounds for functionals in our proofs which are simply provided using Fréchet differentiation. Theorems applying Gâteaux differentiation are not studied in this work. Recent surveys on functional differentiation and its applications in statistics are given by Reeds (1976), Serfling (1980), and Huber (1981).

Definition 3.3.1. *Let D be the space of finite linear combinations of distribution functions. A functional $T_k(F; G_1, \ldots, G_k)$ with F fixed which maps D^k into $\Re$ will be said to be* **k-linear** *if*

$$T_k(F; G_1, \ldots, G_k) = \int \cdots \int h_k(F; x_1, \ldots, x_k) \prod_{i=1}^{k} dG_i(x_i)$$

for some real valued $h_k(F; x_1, \ldots, x_k)$ which is symmetric in $x_1, x_2, \ldots, x_k$. We let $T_k(F; G) \equiv T_k(F; G, \ldots, G)$. The function h_k is said to be the **kernel** *of T_k.*

The function h_1 is usually referred to as the **influence curve** in the literature of statistics.

Definition 3.3.2. *Let $\| \cdot \|$ be a norm on $\mathcal{D}$. Suppose T is a real valued functional defined on $\mathcal{F} \subseteq \mathcal{D}$ where $\mathcal{F}$ contains a neighborhood of $F \in \mathcal{D}$; i.e., there exists $\delta > 0$ such that $G \in \mathcal{D}$ and $\| G \| < \delta$ implies $F + G \in \mathcal{F}$. Let k be a positive integer. Suppose $T_j(F; G_1, \ldots, G_j)$ is a functional defined for $G_i \in \mathcal{D}$, $1 \leq i \leq j$, which is j-linear, $1 \leq j \leq k$. Let $T_0(F; G) \equiv T(F)$. If, for $0 \leq i \leq k$,*

$$\frac{R_i(F;G)}{\| G \|^i} \equiv \frac{T(F+G) - \sum_{j=0}^{i} T_j(F;G)/j!}{\| G \|^i}$$

goes to zero as $\| G \|$ goes to zero then T is said to be **k times Fréchet differentiable** *with respect to the norm $\| \cdot \|$ at F. Furthermore, $T_j(F; \cdot)$ is called the* **j$^{\text{th}}$ Fréchet differential** *of T at F, $0 \leq j \leq k$.*

The usual candidate for $T_j(F; G)$ is

$$T_j(F;G) = \frac{d^j}{dh^j} T(F + hG) \,|_{h=0} \,. \tag{3.3.1}$$

We will not say much about how to find T_j and h_j; examples for M- and L-estimates are given in chapter 4.

In the case it considers, the conditions of the above definition are slightly weaker than those of the standard definition of Fréchet differentiability given by Reeds (1976), p. 151. It would be more appropriate to say that if T satisfies the conditions of definition 3.3.2, then T has a k^{th} order Taylor expansion about F with remainder $o(\| G \|^k)$. For the sake of brevity we do not do this. Because the requirements are weaker, results which assume definition 3.3.2 also hold if the standard definition of Fréchet differentiability is assumed instead. However, we do not need to show that the additional conditions of the latter definition hold to apply our results. We obtain the standard definition of Fréchet differentiation by adding additional assumptions to definition 3.3.2. First, we must have that for some $\epsilon > 0$ and any $H \in \mathcal{D}$ with $\| H - F \| < \epsilon$ the functional T is k times differentiable at H by definition 3.3.2. We must also have that $T_j(H; G_1, \ldots, G_j) - T_j(F; G_1, \ldots, G_j) \to 0$ as $\| H - F \| \to 0$, and that $T_j(H; G_1, \ldots, G_j)$ is uniformly bounded for $\| H - F \| < \epsilon$ if $G_i \in \mathcal{D}$ and $\| G_i \| < 1$, $1 \leq i \leq j \leq k$. We have not shown that any 'interesting' functionals are differentiable by definition 3.3.2 but are not differentiable by the standard definition. The proofs of differentiability given in chapter 4 do not appear to have trivial extensions to show that the conditions of the latter definition hold.

Serfling (1980), p. 217, gives a definition of first order differentiability comparable to definition 3.3.2. He requires only that $T_1(F; \cdot)$ be defined on the space of differences of distribution functions rather than on the linear space generated by distribution functions. This addition to the domain of definition is found to be useful in lemmas 3.4.4 and 3.5.1 below. We can prove the same

results given here if T is only defined for distribution functions by requiring that the results of these lemmas hold. The given definition allows us to develop a slightly more pleasing theory and does not really cost us anything in terms of our particular applications.

Fréchet differentiation is only one type of **functional differentiation**. Other notions of functional differentiation which are useful in statistics are referred to as Gâteaux and compact differentiation. These definitions differ in their requirements on the remainder term, with Fréchet differentiation having the strongest requirement.

See Reeds (1976) for more discussion on many of the above matters.

Definition 3.3.3. *Assume $X_1, X_2, \ldots$ are iid F and let F_n denote the empirical cdf of the first n observations. Suppose the domain of definition of T includes all empirical cdf's. Then the random variable $T(F_n)$ will be referred to as a* **functional statistic**. *The expansion*

$$T(F_n) = \sum_{j=0}^{k} T_j(F; F_n - F)/j! + R_k(F; F_n - F) = \sum_{j=0}^{k} T_{j,n}/j! + R_{k,n}$$

will be referred to as the (k^{th} order) **von Mises expansion** *of $T(F_n)$. The random variable $R_{k,n}$ will be referred to as the* **remainder term** *of the expansion.*

The name von Mises expansion derives from the pioneering work of von Mises (1947) in the application of functional differentiation to statistics. When $F_n - F$ is replaced by an arbitrary $G \in \mathcal{D}$ the expansion is also referred to as a Taylor's series expansion.

The primary applications of functional differentiation in statistics have been to approximate functional statistics using von Mises expansions, and then to extend results for these approximations (which are usually easy to obtain) to the functional statistic. Some typical results obtained are extensions of the central limit theorem, of the law of the iterated logarithm, of the theory of Edgeworth expansions, and of the Berry-Esséen theorem (see Reeds (1976) and Serfling (1980)). As an example of an application which we will use we consider the following central limit theorem which is very similar to results given by Boos and Serfling (1980) and Serfling (1980).

Theorem 3.3.4. *Suppose F is a cdf and T is defined on $\mathcal{F}$ which contains F and all empirical cdf's. For $G \in \mathcal{D}$ let $\| G \|_\infty = \sup_{-\infty < x < \infty} | G(x) |$. Suppose that T has a Fréchet differential T_1 at F with respect to $\| \cdot \|_\infty$ which is not identically 0. If $0 < \sigma^2 = \mathrm{E}[(T_1(F; \delta_{X_1} - F))^2] < \infty$, then $\sqrt{n}(T(F_n) - T(F))$ converges in distribution to $N(0, \sigma^2)$ as $n \to \infty$.*

Proof: Since $\| F_n - F \|_\infty = O_p(n^{-1/2})$ we have from the definition of Fréchet differentiability that

$$T(F_n) - T(F) = T_1(F; F_n - F) + o_p(n^{-1/2}). \tag{3.3.2}$$

Since T_1 is linear in its second argument and $F_n - F = \sum_{i=1}^n (\delta_{X_i} - F)/n$ it follows that

$$T_1(F; F_n - F) = \frac{1}{n} \sum_{i=1}^n T_1(F; \delta_{X_i} - F).$$

By the assumption that $\mathrm{E}[(T_1(F; \delta_{X_1} - F))^2] < \infty$ it follows that $\mathrm{E}[T_1(F; \delta_{X_1} - F)]$ is well defined. From definition 3.3.1 we see that $\mathrm{E}[T_1(F; \delta_{X_1} - F)] = \mathrm{E}[h_1(F; X_1)] - \int h_1(F; x) dF(x) = 0$. Thus by the central limit theorem for iid random variables with finite variance $\sqrt{n} T_1(F; F_n - F)$ converges in distribution to $N(0, \sigma^2)$. The contention now follows from this and (3.3.2). ∎

§3.4 A general moment result.

In this section we will state and prove a result which may be applied to show moment expansions for a wide variety of robust statistics. We begin by stating the basic theorem.

Theorem 3.4.1. *Suppose $X_1, X_2, \ldots$ are iid with cdf F. Let the empirical cdf of X_i, $1 \le i \le n$, be denoted by F_n. Let $X_{1:n}, \ldots, X_{n:n}$ denote the order statistics of X_i, $1 \le i \le n$. Suppose T satisfies the following three conditions:*

i. *T is defined on $\mathcal{F} \subseteq \mathcal{D}$ where $\mathcal{F}$ contains F and all empirical cdf's.*

ii. *T is k times Fréchet differentiable at F with respect to the sup norm $\|\cdot\|_\infty$.*

iii. *There exist constants δ, η, N, and m, all greater than zero, and $\epsilon \in (0, 1/2)$, such that for all $n \ge N$ if $a_n \le \epsilon n$ and $b_n \ge (1-\epsilon)n$ then*

$$| T(F_n) | \le \eta(| X_{a_n:n} | + | X_{b_n:n} | + 2\delta)^m.$$

Suppose F satisfies

iv. $\alpha_+ = \liminf_{x \to \infty} \dfrac{-\log(1 - F(x))}{\log x} > 0$ *and* $\alpha_- = \liminf_{x \to \infty} \dfrac{-\log F(-x)}{\log x} > 0.$

Recall that $T_{j,n} \equiv T_j(F; F_n - F)$ is the j^{th} differential of T at F in the direction $F_n - F$. Under assumptions i–iv, for any positive integer r

$$\mathrm{E}[(T(F_n) - T(F))^r] = \mathrm{E}\left[\left(\sum_{j=1}^k T_{j,n}/j!\right)^r\right] + o(n^{-(r+k-1)/2}) \tag{3.4.1}$$

and

$$(\mathrm{E}[|T(F_n) - T(F)|^r])^{1/r} = \left(\mathrm{E}\left[\left|\sum_{j=1}^k T_{j,n}/j!\right|^r\right]\right)^{1/r} + o(n^{-k/2}). \tag{3.4.2}$$

Remarks.

1) If we replace the inequality of *iii* with

$$X_{a_n:n} - \delta < T(F_n) < X_{b_n:n} + \delta$$

and *iv* does not hold, then $\mathrm{E}[| T(F_n) - T(F) |^r] = \infty$ by theorem 2.2.6. Thus *iv* is a necessary condition in this case.

2) We will argue (not rigorously) in section 3.5 that if T is $k+1$ times Fréchet differentiable then the remainder term in (3.4.1) should be $O(n^{-[\![(r+k+1)/2]\!]})$ where $[\![a]\!]$ denotes the greatest integer less than or equal to a. The general method of approximating $\mathrm{E}[(\sum T_{j,n})^r]$ will also be given in the section 3.5. The remainder of this section consists of a series of lemmas which will be used to prove theorem 3.4.1. The fundamental lemma to be used is the following:

Lemma 3.4.2. *Suppose $X_1, X_2, \ldots$ are iid with cdf F. Let T be a functional which is defined on $\mathcal{F} \subseteq D$ where $\mathcal{F}$ contains F and all empirical cdf's. Suppose also that T satisfies conditions i—iii and F satisfies iv of theorem 3.4.1. Let*

$$R_{k,n} = T(F_n) - T(F) - \sum_{j=1}^{k} T_{j,n}/j!$$

be the remainder of the k^{th} order von Mises expansion. Then for any positive integer r

$$\mathrm{E}[(n^{k/2} \mid R_{k,n} \mid)^r] \to 0 \quad as \quad n \to \infty.$$

We delay the proof of this lemma until some preliminary results have been established. The first two lemmas are the keys to proving our main results. Lemma 3.4.3 was first given by Dvoretsky, Kiefer, and Wolfowitz (1956). It is very useful for obtaining uniform integrability results when using Fréchet differentiation with respect to $\|\cdot\|_\infty$. Lemma 3.4.4 gives bounds for the differentials $T_k(F;\cdot)$ which will be needed.

Lemma 3.4.3. *Let $D_n = \| F_n - F \|_\infty$ be the Kolmogorov-Smirnov statistic after n iid observations from F. Then there exists $c_1 > 0$ such that for all n and x*

$$\mathrm{P}\{\sqrt{n} D_n > x\} \leq c_1 e^{-2x^2}.$$

Lemma 3.4.4. *If T is k times Fréchet differentiable at F with respect to the a norm $\|\cdot\|$ on $\mathcal{D}$, then there exists $A > 0$ such that for any $G \in \mathcal{D}$, $| T_k(F;G) | \le A \| G \|^k$.*

Proof: Let $\delta > 0$. From the differentiability assumption we know that there exists $\epsilon_\delta > 0$ such that if $\| G \| = \epsilon_\delta$ then $| R_k(F;G) | \le \delta \| G \|^k$ and $| R_{k-1}(F;G) | \le \delta \| G \|^{k-1}$. Note that this holds even for $k = 1$. Since $T_k(F;G) = R_{k-1}(F;G) - R_k(F;G)$ this implies that $| T_k(F;G) | \le \delta \| G \|^{k-1} (1 + \| G \|)$.

Now suppose the contention is not true. Then there exist $G_i \in \mathcal{D}$, $i = 1, 2, \ldots$ such that $T_k(F;G_i) > i \| G_i \|^k$. Since for any $a \in \Re$, $G \in \mathcal{D}$ we have $T_k(F;aG_i) = a^k T_k(F;G_i)$ we may assume without loss of generality that $\| G_i \|_\infty = \epsilon_\delta$. Since for i sufficiently large $T_k(F;G_i) > \delta \epsilon_\delta^{k-1}(1 + \epsilon_\delta)$ we have contradicted what we have shown in the paragraph above. ∎

The following definition of uniform integrability and corresponding result are slight variations on those given by Breiman (1968), p. 91. The result here follows from Breiman's result.

Definition 3.4.5. *A sequence of random variables Y_n, $n = 1, 2, \ldots$ will be said to be* **uniformly integrable** *if for any $\epsilon > 0$ there exist A_ϵ and N_ϵ such that for all $n \ge N_\epsilon$,*

$$\mathrm{E}[| Y_n | I\{| Y_n | \ge A_\epsilon\}] < \epsilon.$$

Lemma 3.4.6. *If Y_n converges in distribution to Y as $n \to \infty$ and Y_n, $n = 1, 2, \ldots$, is a uniformly integrable sequence, then $\mathrm{E}[Y_n] \to \mathrm{E}[Y]$ as $n \to \infty$.*

Lemma 3.4.7. *For any positive integer r, $(n^{1/2} D_n)^r$ is uniformly integrable.*

Proof: This result follows from lemma 3.4.3 and definition 3.4.5 after applying integration by parts. ∎

Lemma 3.4.8. *Suppose F is a cdf. Suppose T is defined on $\mathcal{F} \subseteq \mathcal{D}$ where $\mathcal{F}$ contains F and all empirical cdf's. Suppose T is k times Fréchet differentiable at F with respect to the sup norm $\|\cdot\|_\infty$. Recall $T_{k,n} \equiv T_k(F; F_n - F)$ is the k^{th} differential of $T(F)$ evaluated at $F_n - F$. If r is a positive integer, then $(n^{k/2} T_{k,n})^r$, $n = 1, 2, \ldots$, is uniformly integrable.*

Proof: From lemma 3.4.4 we know that there exists $A > 0$ such that $| T_{k,n} | \le A D_n^k$. From lemma 3.4.7 we know that $(n^{k/2} D_n^k)^r$ is uniformly integrable and the contention follows. ∎

Lemma 3.4.9. *Suppose r is a positive integer, and X_n^r and Y_n^r are uniformly integrable. Then so is $(X_n+Y_n)^r$.*

Proof: Since $|X_n+Y_n|^r \le 2^r(\max(|X_n|,|Y_n|))^r$ it follows that

$$\begin{aligned}\mathrm{E}[|X_n+Y_n|^r I\{|X_n+Y_n|^r \ge A\}] \le &\mathrm{E}[2^r |X_n|^r I\{|X_n|^r \ge A2^{-r}\}] \\ &+\mathrm{E}[2^r |Y_n|^r I\{|Y_n|^r \ge A2^{-r}\}].\end{aligned}$$

Under the hypothesis, for any $\delta > 0$ there exists A such that the right hand side of the above is less than δ for n sufficiently large. ∎

Proof of lemma 3.4.2: Since the remainder $R_{k,n} = o(\| F_n - F \|_\infty^k)$ and $\| F_n - F \|_\infty = O_p(n^{-1/2})$ it follows that $n^{k/2}R_{k,n}$ goes in probability to 0. By lemma 3.4.6, the only thing that remains to be shown is that $(n^{k/2} | R_{k,n} |)^r$ is uniformly integrable. We let $0 < \gamma < 1/2$, $c > 0$, and let r be a positive integer. We will use the decomposition

$$(n^{k/2} | R_{k,n} |)^r = (n^{k/2} | R_{k,n} |)^r I\{D_n \le cn^{-\gamma}\} + (n^{k/2} | R_{k,n} |)^r I\{D_n > cn^{-\gamma}\}. \qquad (3.4.3)$$

The first term of (3.4.3) will be considered first. Choose $\delta > 0$. From the definition of Fréchet differentiability we may choose ϵ_δ so that for all G with $\| G \|_\infty \le \epsilon_\delta$, $F+G \in \mathcal{F}$ and $| R_k(F;G) | < \| G \|_\infty^k \delta$. Let $N \ge (c/\epsilon_\delta)^{1/\gamma}$. If $n \ge N$ then $cn^{-\gamma} \le \epsilon_\delta$ and

$$n^{k/2} | R_{k,n} | I\{D_n \le cn^{-\gamma}\} \le \delta n^{k/2} D_n^k.$$

From lemma 3.4.7 it follows that $(n^{k/2} | R_{k,n} |)^r I\{D_n \le cn^{-\gamma}\}$ is uniformly integrable.

Now we consider the second term of (3.4.3). Since $R_{k,n} = T(F_n) - T(F) - \sum_{j=1}^k T_{j,n}/j!$ it is sufficient (from lemma 3.4.9) to show the uniform integrability of $(n^{k/2} | T(F_n) |)^r I\{D_n > cn^{-\gamma}\}$, $(n^{k/2} | T(F) |)^r I\{D_n > cn^{-\gamma}\}$ and $(n^{k/2} | T_{j,n} |)^r I\{D_n > cn^{-\gamma}\}$, $j = 1,\ldots,k$.

By lemma 3.4.3 $(n^{k/2} | T(F) |)^r I\{D_n > cn^{-\gamma}\}$ is uniformly integrable.

From lemma 3.4.4 we know that $| T_{j,n} | < AD_n^j$ for some $A > 0$. Applying lemma 3.4.3 and integration by parts we see that

$$\begin{aligned}\mathrm{E}[(n^{k/2} | T_{j,n} |)^r I\{D_n > cn^{-\gamma}\}] &\le A^r n^{r(k-j)/2}\mathrm{E}[(n^{1/2}D_n)^{jr} I\{n^{1/2}D_n > cn^{1/2-\gamma}\}] \\ &\le A^r n^{r(k-j)/2}\Big(c_1 \exp\left(-2c^2 n^{1-2\gamma}\right) c^{jr} n^{jr(1/2-\gamma)} \\ &\qquad + \int_{cn^{1/2-\gamma}}^\infty jr x^{jr-1} c_1 e^{-2x^2}\, dx\Big).\end{aligned}$$

The first term inside the brackets times the factor outside the brackets goes to zero as $n \to \infty$. Letting $y = 2x^2$ we can find positive constants c_2 and c_3 such that the second term inside the brackets times the factor outside is less than or equal to

$$c_2 n^{r(k-j)/2} \int_{c_3 n^{1-2\gamma}}^{\infty} y^{(jr-2)/2} e^{-y} dy.$$

If n is sufficiently large, a is a positive integer greater than $(jr-2)/2$, and $\epsilon = 1 - 2\gamma > 0$, then this is less than

$$c_2 n^{r(k-j)/2} a! \exp(-c_3 n^{\epsilon}) \sum_{j=0}^{a} (c_3 n^{\epsilon})^j / j!$$

which goes to zero as $n \to \infty$. Thus $\mathrm{E}[(n^{k/2} \mid T_{j,n} \mid)^r I\{D_n > cn^{-\gamma}\}]$ goes to 0 as $n \to \infty$, $j = 1, 2, \ldots, k$.

Finally we consider $(n^{k/2} \mid T(F_n) \mid)^r \; I\{D_n > cn^{-\gamma}\}$. We let ϵ and δ be as in assumption *iii* of theorem 3.4.1. Let $a_n = [\![n\epsilon]\!]$ and $b_n = [\![n(1-\epsilon)]\!] + 1$. Let $\beta \in (0, 1/4)$ and consider

$$\begin{aligned}
&\left(n^{k/2}(\mid X_{b_n:n} \mid + \delta)\right)^{rm} I\{D_n > cn^{-\gamma}\} \\
&\qquad = \left(n^{k/2}(\mid X_{b_n:n} \mid + \delta)\right)^{rm} I\{D_n > cn^{-\gamma}\} I\{\mid X_{b_n:n} \mid > F^{-1}(1 - \epsilon + cn^{-\beta})\} \\
&\qquad\quad + \left(n^{k/2}(\mid X_{b_n:n} \mid + \delta)\right)^{rm} I\{D_n > cn^{-\gamma}\} I\{\mid X_{b_n:n} \mid \leq F^{-1}(1 - \epsilon + cn^{-\beta})\}
\end{aligned}$$

The first term is uniformly integrable by lemma 2.3.2. The second term is bounded by $\left(n^{k/2}(F^{-1}(1 - \epsilon + cn^{-\beta}) + \delta)\right)^{rm} \mathrm{P}\{D_n > cn^{-\gamma}\}$. It goes to zero by lemma 3.4.3 since $\gamma < 1/2$. By lemma 3.4.9, $\left(n^{k/2}(\mid X_{b_n:n} \mid + \delta)\right)^{rm} \; I\{D_n > cn^{-\gamma}\}$ is uniformly integrable. Similarly $\left(n^{k/2}(\mid X_{a_n:n} \mid + \delta)\right)^{rm} \; I\{D_n > cn^{-\gamma}\}$ is uniformly integrable. By lemma 3.4.9 it follows that $\left(n^{k/2}(\mid X_{a_n:n} \mid + \mid X_{b_n:n} \mid + 2\delta)\right)^{rm} I\{D_n > cn^{-\gamma}\}$ is uniformly integrable. By assumption *iii* it follows that $\left(n^{k/2} T(F_n)\right)^r I\{D_n > cn^{-\gamma}\}$ is a uniformly integrable sequence. ∎

Finally we prove our main theorem.

Proof of theorem 3.4.1. Since

$$T(F_n) - T(F) = \sum_{j=1}^{k} T_{j,n}/j! + R_{k,n} \tag{3.4.4}$$

(3.4.2) follows from Minkowski's inequality and lemma 3.4.2.

Using (3.4.4) and expanding we have

$$\mathrm{E}[(T(F_n) - T(F))^r] = \sum_{i=0}^{r} \binom{r}{i} \mathrm{E}\left[R_{k,n}^i \left(\sum_{j=1}^{k} T_{j,n}/j! \right)^{r-i} \right].$$

For $1 \leq i \leq r$ it follows from Hölder's inequality that

$$\mathrm{E}\left[\left| R^i_{k,n}\left(\sum_{j=1}^{k} T_{j,n}/j!\right)^{r-i}\right|\right] \leq (\mathrm{E}[\| R_{k,n} \|^r])^{i/r}\left(\mathrm{E}\left[\left|\sum_{j=1}^{k} T_{j,n}/j!\right|^r\right]\right)^{(r-i)/r}.$$

From lemma 3.4.8 it follows that $\mathrm{E}[\| T_{j,n} \|^r]^{1/r} = O(n^{-j/2})$ for $j = 1, 2, \ldots, k$. From this and lemma 3.4.2 it follows that

$$(\mathrm{E}[\| R_{k,n} \|^r])^{i/r}\left(\mathrm{E}\left[\left|\sum_{j=1}^{k} T_{j,n}/j!\right|^r\right]\right)^{(r-i)/r} = o(n^{-(ik+r-i)/2})$$

and since $ik + r - i \geq r + k - 1$ (3.4.1) follows. ∎

§3.5 Calculating moment approximations.

In the previous section we have proven convergence of some moment approximations. We will now present some results on functional differentiation and von Mises expansions which are useful in applying this theory to actually calculate moment approximations. We also give formulas for bias, and first and second order mean squared error approximations in terms of the first three functional derivatives of a statistic. These formulas will be applied for L- and M-estimates in the next chapter.

We assume throughout the following that $X_1, X_2, \ldots$ is a sequence of iid random variables with cdf F and that T is a functional defined in a neighborhood of F and for all empirical cdf's which is k times Fréchet differentiable at F with respect to $\|\cdot\|_\infty$. It is a simple consequence of definition 3.3.1 that

$$T_k(F; G_1 + G_1^*, G_2, \ldots, G_k) = T_k(F; G_1, G_2, \ldots, G_k) + T_k(F; G_1^*, G_2, \ldots, G_k) \tag{3.5.1}$$

and

$$T_k(F; aG_1, G_2, \ldots, G_k) = aT_k(F; G_1, G_2, \ldots, G_k). \tag{3.5.2}$$

Lemma 3.5.1 is used in the proof of lemma 3.5.2 which shows that the kernels of these Fréchet differentials are bounded when differentiation is done with respect to the sup norm.

Lemma 3.5.1. *Suppose k is a positive integer and T is a functional defined on $\mathcal{F} \in \mathcal{D}$ which is k times Fréchet differentiable at F with respect to a norm $\| \cdot \|$ on $\mathcal{D}$. Then for $j = 1, 2, \ldots, k$, $T_j(F; G_1 \ldots, G_j)$ goes to zero uniformly as $\max_{1 \leq i \leq j} \| G_i \|$ goes to zero.*

Proof: Fix j and m such that $1 < m < j \leq k$. Assume that if $\{G_1, \ldots, G_k\}$ contains exactly $m-1$ distinct elements then $T_j(F; G_1, \ldots, G_j)$ goes to zero uniformly as $\max_{1 \leq i \leq j} \| G_i \|$ goes to zero. Now assume that $\{G_1, \ldots, G_k\}$ contains exactly m distinct elements and that $G_1 \neq G_2$. By (3.5.1) and the symmetry of $T_j(F; \cdot)$ it follows that

$$T_j(F; G_1, \ldots, G_j) = (1/2)\left(T_j(F; G_1 + G_2, G_1 + G_2, G_3, \ldots, G_j) - \sum_{i=1}^{2} T_j(F; G_i, G_i, G_3, \ldots, G_j). \right)$$

Since $T_j(F; \cdot)$ on the right hand side of this equation has $m - 1$ distinct arguments in each case, it follows that they all go to zero uniformly as $\max_{1 \leq i \leq j} \| G_i \|$ goes to zero. The contention follows for $m = 1$ and arbitrary j, $1 \leq j \leq k$, by lemma 3.4.4. For arbitrary j, $1 < j \leq k$, the result follows for $1 < m \leq j$ by induction on m. ∎

Lemma 3.5.2. *Let k be a positive integer. Suppose T is k times Fréchet differentiable at F with respect to $\| \cdot \|_\infty$. Then the kernel h_k of T_k is bounded.*

Proof: The proof uses contraposition. Suppose $h_k(F; \cdot)$ is not bounded. Then there exist x_{im}, $1 \leq i \leq k$, such that $h_k(F; x_{1m}, \ldots, x_{km}) > m^{k+1}$, $m = 1, 2, \ldots$. Let $G_{im} = \delta_{x_{im}}/m$. This implies that $T_k(F; G_{1m}, \ldots, G_{km}) > m$ and that $\max_{1 \leq j \leq k} \| G_{jm} \|_\infty = 1/m$ goes to zero as $m \to \infty$. This contradicts lemma 3.5.1. ∎

We let

$$Z_i(x) = \delta_{X_i}(x) - F(x). \tag{3.5.3}$$

Since

$$F_n - F = n^{-1} \sum_{i=1}^{n} Z_i \tag{3.5.4}$$

it follows from (3.5.1) and (3.5.2) that

$$T_{k,n} = T_k(F; F_n - F) = n^{-k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} T_k(F; Z_{i_1}, \ldots, Z_{i_k}). \tag{3.5.5}$$

Since h_k is bounded we know that $T_k(F; Z_{i_1}, \ldots, Z_{i_k})$ is a bounded random variable. If T has a k^{th} differential and r is a positive integer then by using (3.5.5) and expanding we find

$$\mathrm{E}[(\sum_{j=1}^{k} T_{j,n}/j!)^r] = \sum \mathrm{E}[\prod_{l=1}^{r} n^{-j(l)} T_{j(l)}(F; Z_{i_{l1}}, \ldots, Z_{i_{lj(l)}})/j(l)!], \tag{3.5.6}$$

where the sum is appropriately defined (explicit examples of (3.5.6) will be derived more carefully in the proof of theorem 3.5.3). Symmetry arguments tell us that we need only compute a 'few' of the terms of the right hand side of (3.5.6) to get $\mathrm{E}[(\sum_{j=1}^{k} T_{j,n})^r]$. The following lemma shows that even most of these terms are zero.

Lemma 3.5.3. *Suppose at least one of the indices i_{lm} occurs exactly once in*

$$\prod_{l=1}^{r} T_{j(l)}(F; Z_{i_{l1}}, \ldots, Z_{i_{lj(l)}})/j(l)!.$$

Then

$$\mathrm{E}[\prod_{l=1}^{r} T_{j(l)}(F; Z_{i_{l1}}, \ldots, Z_{i_{lj(l)}})/j(l)!] = 0.$$

Proof: Without loss of generality we assume $i_{1,1} = 1$, and if $(l,m) \neq (1,1)$ then $i_{lm} \neq i_{1,1}$. Let $j = j(1)$. For a given i we condition on $Z_{i_{lm}} = z_{lm}$ for $(l,m) \neq (1,1)$. It will suffice to show

$$\mathrm{E}[T_j(F; Z_1, z_{12}, \ldots, z_{1j})] = 0.$$

By definition

$$\mathrm{E}[T_j(F; Z_1, z_{12}, \ldots, z_{1j})] = \mathrm{E}[\int \cdots \int h_j(F; x_1, \ldots, x_j) dZ_1(x_1) \prod_{l=2}^{j} dz_{1l}(x_l)].$$

Since h_j is bounded by lemma 3.5.2 and all of the measures involved are bounded we may switch the order of integration to obtain

$$\mathrm{E}[T_j(F; Z_1, z_{i_{12}}, \ldots, z_{i_{1j}})] = \int \cdots \int \mathrm{E}[\int h_j(F; x_1, \ldots, x_j) dZ_1(x_1)] \prod_{l=2}^{j} dz_{1l}(x_l).$$

But

$$\mathrm{E}[\int h_j(F; x_1, \ldots, x_j) dZ_1(x_1)] = \mathrm{E}[h_j(F; X_1, x_2, \ldots, x_j)] - \int h_j(F; x_1, \ldots, x_j) dF(x_1) = 0$$

and the contention follows. ∎

Using (3.5.6) and lemma 3.5.3 we can show by counting the number of terms of various types that for $k \geq 1$

$$\mathrm{E}[(T(F_n) - T(F))^r] = \mathrm{E}[(\sum_{j=1}^{k} T_{j,n}/j!)^r] + O(n^{-[\![(r+k+1)/2]\!]}) \tag{3.5.7}$$

if T is $k+1$ times Fréchet differentiable. This was suggested after the statement of theorem 3.4.1 and will be shown for some particular cases in theorem 3.5.4.

We are now prepared to give formulas for first and second order mean and mean squared error approximations.

Theorem 3.5.4. *Suppose $X_1, X_2, \ldots$ are iid with cdf F and that T is a functional defined in a neighborhood of F and for all empirical cdf's. For the remainder of this proposition if we say T is k times Fréchet differentiable, we mean at F with respect to $\|\cdot\|_\infty$; also implicit in this statement will be that the approximation*

$$E[(T(F_n) - T(F))^r] = E[(\sum_{j=1}^{k} T_{j,n}/j!)^r] + o(n^{-(r+k-1)/2})$$

is valid. Define $Z_i(x) = \delta_{X_i}(x) - F(x)$ as before. If T is (once) Fréchet differentiable then

$$E[T(F_n) - T(F)] = o(1/\sqrt{n}) \tag{3.5.8}$$

and

$$E[(T(F_n) - T(F))^2] = \frac{1}{n}E[(T_1(F; Z_1))^2] + o(1/n). \tag{3.5.9}$$

If T is two times Fréchet differentiable then

$$E[T(F_n) - T(F)] = \frac{1}{2n}E[T_2(F; Z_1, Z_1)] + o(1/n). \tag{3.5.10}$$

If T is three times Fréchet differentiable then

$$\begin{aligned} E[(T(F_n) - T(F))^2] =& \frac{1}{n}E[(T_1(F; Z_1))^2] + \frac{1}{n^2}E[T_1(F; Z_1)T_2(F; Z_1, Z_1)] \\ &+ \frac{1}{2n^2}E[(T_2(F; Z_1, Z_2))^2] + \frac{1}{4n^2}(E[T_2(F; Z_1, Z_1)])^2 \\ &+ \frac{1}{n^2}E[T_1(F; Z_1)T_3(F; Z_1, Z_2, Z_2)] + o(1/n^2). \end{aligned} \tag{3.5.11}$$

Proof: Assume T is one time Fréchet differentiable. By assumption

$$E[T(F_n) - T(F)] = E[T_1(F; F_n - F)] + o(1/\sqrt{n}).$$

From (3.5.1)—(3.5.5) we have

$$E[T_1(F; F_n - F)] = E[n^{-1} \sum_{i=1}^{n} T_1(F; Z_i)].$$

Since $E[T_1(F; Z_i)]$ exists and is finite we may interchange the order of integration and summation. From lemma 3.5.3 we have $E[T_1(F; Z_i)] = 0$ and (3.5.8) follows.

Similar arguments justify the following calculation:

$$\begin{aligned} E[(T(F_n) - T(F))^2] &= E[(T_1(F; F_n - F))^2] + o(1/n) \\ &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[T_1(F; Z_i)T_1(F; Z_j)] + o(1/n). \end{aligned}$$

By symmetry we have $\mathrm{E}[(T_1(F;Z_i))^2] = \mathrm{E}[(T_1(F;Z_1)^2]$. By lemma 3.5.3 we have $\mathrm{E}[T_1(F;Z_i)T_1(F;Z_j)] = 0$ for $j \neq i$, $1 \leq i \leq n$. Thus we have shown (3.5.9).

To obtain (3.5.10) we use a second order approximation:

$$\begin{aligned}\mathrm{E}[T(F_n) - T(F)] &= \mathrm{E}[T_1(F;F_n - F) + (1/2)T_2(F;F_n - F)] + o(1/n)\\ &= (2n)^{-1}\mathrm{E}[T_2(F;Z_1,Z_1)] + o(1/n).\end{aligned}$$

To obtain (3.5.11) we use a third order approximation:

$$\begin{aligned}\mathrm{E}[(T(F_n) - T(F))^2] &= \mathrm{E}[(T_1(F;F_n - F) + (1/2)T_2(F;F_n - F) + (1/6)T_3(F;F_n - F))^2] + o(1/n^2)\\ &= \mathrm{E}[(T_1(F;F_n - F))^2] + \mathrm{E}[T_1(F;F_n - F)T_2(F;F_n - F)]\\ &\quad + (1/3)\mathrm{E}[T_1(F;F_n - F)T_3(F;F_n - F)] + (1/4)\mathrm{E}[(T_2(F;F_n - F))^2] + o(1/n^2).\end{aligned}$$

The terms which 'disappear' in the last line are $o(1/n^2)$ from lemma 3.4.8 and Hölder's inequality. We have already derived $\mathrm{E}[T_1(F;F_n - F))^2]$ above. We continue with the other terms.

$$\begin{aligned}\mathrm{E}[T_1(F;F_n - F)T_2(F;F_n - F)] &= n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{E}[T_1(F;Z_i)T_2(F;Z_j,Z_k)]\\ &= n^{-2}\mathrm{E}[T_1(F;Z_1)T_2(F;Z_1,Z_1)].\end{aligned}$$

$$\begin{aligned}\mathrm{E}[T_1(F;F_n - F)T_3(F;F_n - F)] &= n^{-4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{E}[T_1(F;Z_i)T_3(F;Z_j,Z_k,Z_l)]\\ &= n^{-4}\binom{n}{2}\cdot 2\cdot 3\cdot \mathrm{E}[T_1(F;Z_1)T_2(F;Z_1,Z_2,Z_2)] + o(1/n^2).\end{aligned}$$

$$\mathrm{E}[(T_2(F;F_n - F))^2] = n^{-4}\left(n(n-1)(\mathrm{E}[T_2(F;Z_1,Z_1)])^2 + \binom{n}{2}\cdot 4\cdot \mathrm{E}[(T_2(F;Z_1,Z_2))^2]\right).$$

Equation (3.5.11) now follows from these calculations. ∎

Chapter 4

L- and M-estimates

§4.1 Introduction.

In this chapter we will apply the theory of the previous chapter to derive expansions for moments of many L- and M-estimates. In particular, we will give first and second order approximations for the mean and mean squared error in each case. Applications of these approximations will be studied in the next chapter.

We now note some of the limitations of the theory of chapter 3 in applications to L- and M-estimates. Most of the limitations should not be of great concern to those wishing to consider robust statistics.

The first limitation is that the influence function (kernel of the first differential) must be bounded. This is not an important limitation for robustness since it implies that changing a small proportion of the observations, in general, changes the value of the statistic by at most a limited amount. For L-estimates this limitation means we must exclude a positive proportion of the extreme observations from the calculation of the estimate.

For M-estimates we have the requirement that the influence function be non-decreasing as well as bounded. We need this to give the quantile bound required by theorem 3.4.1. This eliminates M-estimates with redescending influence curves, which is an unfortunate limitation. Another limitation of the present treatment is that we have not considered simultaneous estimation of location and scale for M-estimates. Eynon (1982) uses some of the results given here to treat this case.

Finally, the condition of Fréchet differentiability imposes 'strong smoothness' conditions on the estimators to be considered. For instance, quantiles are not Fréchet differentiable, and the

M-estimate referred to as Huber's proposal number one is only once Fréchet differentiable. Examples of functionals which are not Fréchet differentiable and efforts to patch up our theory in these cases are discussed in section 5.4. We note difficulties in showing that the standard definition of Fréchet differentiability holds in cases where we show our definition of Fréchet differentiability holds after the proofs of proposition 4.2.5 and lemma 4.3.5.

§4.2 Theory for L-estimates.

We will now apply the theory of chapter 3 to develop a theory of moment convergence for L-estimates. The results on moment approximation contained in this section are summarized by (4.2.1), lemma 4.2.4 and proposition 4.2.6. The L-estimates we will consider are defined for any cdf F by

$$T(F) = \int_0^1 J(u)F^{-1}(u)du \tag{4.2.1}$$

when the integral is well defined and J is a real valued function. We do not use the more general version

$$T(F) = \int_0^1 J(u)F^{-1}(u)du + \sum_{i=1}^{m} a_i X_{[\![c_i n]\!]+1:n} \tag{4.2.2}$$

since then T would not be Fréchet differentiable; this is discussed in section 5.4. Recall that a result on the moments of $\sum_{i=1}^{m} a_i X_{[\![c_i n]\!]+1:n}$ was given in proposition 2.3.11.

Note that the functional of (4.2.1) is not defined on all of $\mathcal{D}$ since not all elements of $\mathcal{D}$ have an inverse. Thus before showing the Fréchet differentiability of the functional defining an L-estimate, we must extend the definition to $\mathcal{D}$. The concept of bounded variation and some simple properties related to it will be needed for this definition and for the proof of Fréchet differentiability for both L- and M-estimates.

Definition 4.2.1. *A function g is said to have* **bounded variation** *on $[a, b]$ if*

$$V_{ab} \equiv \sup\left\{\sum_{i=1}^{n} |\, g(x_i) - g(x_{i-1}) \,| \quad : n < \infty \text{ and } a \le x_0 < \cdots < x_n \le b\right\} < \infty.$$

If

$$\|\, g \,\|_{TV} \equiv \lim_{\substack{a \to -\infty \\ b \to \infty}} V_{ab} < \infty,$$

we say that g has **bounded total variation.**

If $\| g \|_{TV} < \infty$ then g is bounded and may be written as the difference of two bounded, monotone non-decreasing functions. Thus for any function of bounded variation there is a corresponding finite signed measure. When we write an integral with a differential element dg where g is a function of bounded variation this will mean integration is to be done with respect to the finite signed measure corresponding to g. If g is continuous and has bounded variation on $[a, b]$ then it is also uniformly continuous there. The following lemma is a special case of a result given by Rudin (1964), p. 122.

Lemma 4.2.2. *Suppose f and g are real-valued functions of bounded variation on $[a, b]$ and f is also continuous. Then*

$$\int_a^b f dg = f(b)g(b) - f(a)g(a) - \int_a^b g df.$$

Definition 4.2.3. *Let J be a function of bounded variation on $[0,1]$ which is 0 outside of $[0,1]$. For any $G \in \mathcal{D}$ where the integral is well defined we let*

$$T(G) = \int_{-\infty}^{\infty} x d \int_0^{G(x)} J(u) du.$$

If F_n is an empirical cdf generated by a set of n iid random variables, then $T(F_n)$ is said to be an **L-estimate.**

We will be interested in the case in which J is zero outside of some interval $[\delta_1, 1 - \delta_2]$ where $0 < \delta_1 < 1 - \delta_2 < 1$. In this case we now show that if F is a cdf then $T(F)$ of definition 4.2.3 has the same value as $T(F)$ of $\int_0^1 J(u) F^{-1}(u) du$.

Lemma 4.2.4. *Suppose F is a cdf and J is a function of bounded variation on $[0,1]$ which is zero outside of some interval $[\delta_1, 1 - \delta_2]$ where $0 < \delta_1 < 1 - \delta_2 < 1$. Then*

$$\int_0^1 J(u) F^{-1}(u) du = \int_{-\infty}^{\infty} x d\left(\int_0^{F(x)} J(u) du \right).$$

Proof: We begin by rewriting the single integral of the lemma as a sum of two double integrals.

$$\int_0^1 J(u) F^{-1}(u) du = -\int_0^{F(0)} J(u) \int_{F^{-1}(u)}^0 dx du + \int_{F(0)}^1 J(u) \int_0^{F^{-1}(u)} dx du. \tag{4.2.3}$$

We will not show here that $\int_{F(x)}^{1-\delta_2} J(u) du$ is a function of bounded variation; the argument is similar to one given below in the proof of proposition 4.2.5. Let a be such that $F(a) < \delta_1$ and $| a | < \infty$. If $a < 0$ the first of these terms may be rewritten using first Fubini's theorem and then lemma 4.2.2

as

$$-\int_{F(a)}^{F(0)} J(u) \int_{F^{-1}(u)}^{0} dx du = -\int_{a}^{0} \int_{F(a)}^{F(x)} J(u) du dx$$
$$= -x \int_{F(a)}^{F(x)} J(u) du \Big|_{a}^{0} + \int_{a}^{0} x d\left(\int_{F(a)}^{F(x)} J(u) du \right)$$
$$= \int_{a}^{0} x d\left(\int_{0}^{F(x)} J(u) du \right).$$

The result follows by using (4.2.3), this expression, and the analogous expression for the second term of (4.2.3). ∎

For the following proposition recall the definition of a kernel (definition 3.3.1) and our definition of Fréchet differentiability (definition 3.3.2) which has weaker conditions that the standard definition. Our proof is an extension of the analysis of Boos (1979); his proof also appears in Serfling (1980), pp. 281-282. Boos' proof is for the first differential. Note that when we say J is continuous *a.e.* with respect to F^{-1} for some cdf F, we mean with respect to the measure corresponding to the monotone function F^{-1}.

Proposition 4.2.5. *Let J be defined on $[0,1]$ with $J(u) = 0$ for $u < \delta_1$ and $u > 1 - \delta_2$ where $0 < \delta_1 < 1 - \delta_2 < 1$. Let $\mathcal{F}$ be the elements of $\mathcal{D}$ for which*

$$T(G) = \int_{-\infty}^{\infty} x d\left(\int_{0}^{G(x)} J(u) du \right)$$

is well defined. Suppose $F \in \mathcal{F}$ is a cdf. Let k be a positive integer. Let $J^{(j)}$ denote the j^{th} derivative of J where it exists, $j = 1, 2, \ldots, k$. Let $J^{(0)} \equiv J$. Assume that $J^{(j)}$, $j = 0, 1, \ldots, k-2$, are bounded and absolutely continuous. Assume that $J^{(k-1)}$ is bounded and continuous a.e. with respect to F^{-1}. Then T is k times Fréchet differentiable at F with respect to $\|\cdot\|_\infty$. For $j = 1, \ldots, k$ the kernel of the j^{th} Fréchet differential of T at F with respect to $\|\cdot\|_\infty$ is

$$h_j(F; x_1, \ldots, x_j) = -\int_{\max x_i}^{\infty} J^{(j-1)}(F(y)) dy. \tag{4.2.4}$$

We delay the proof of this proposition. For the remainder of the section when limits of integration are not given explicitly, they will always be $\pm\infty$.

Proposition 4.2.6. *Let F be a cdf such that*

$$0 < \liminf_{x \to \infty} \frac{-\log(1 - F(x))}{\log x}, \quad 0 < \liminf_{x \to \infty} \frac{-\log F(-x)}{\log x}.$$

Let $J(u)$ be defined on $[0,1]$ with $J(u)=0$ for $u<\delta_1$ and $u>1-\delta_2$ where $0<\delta_1<1-\delta_2<1$. Let $\mathcal{F}$ be the elements of $\mathcal{D}$ for which

$$T(G)=\int_{-\infty}^{\infty} x d\left(\int_0^{G(x)} J(u)du\right)$$

is well defined. Let k be a positive integer. Assume that $J^{(j)}$, $j=0,2,\ldots,k-2$, are bounded and absolutely continuous. Assume that $J^{(k-1)}$ is bounded and continuous a.e. with respect to F^{-1}. Let $T_{j,n}$ denote the j^{th} Fréchet differential $T_j(F;F_n-F)$ of T at F, $j=1,2,\ldots,k$. Then for any positive integer r

$$\mathrm{E}[(T(F_n)-T(F))^r]=\mathrm{E}\left[\left(\sum_{j=1}^{k} T_{j,n}/j!\right)^r\right]+o(n^{-(r+k-1)/2}). \tag{4.2.5}$$

If $k\geq 1$ then

$$\mathrm{E}[T(F_n)-T(F)]=o(1/\sqrt{n}) \tag{4.2.6}$$

and

$$\mathrm{E}[(T(F_n)-T(F))^2]=\frac{1}{n}\int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)\prod_{i=1}^{2} J(F(x_i))dx_i+o(1/n). \tag{4.2.7}$$

If $k\geq 2$ then

$$\mathrm{E}[T(F_n)-T(F)]=\frac{1}{2n}\int F(x)(1-F(x))J^{(1)}(F(x))dx+o(1/n). \tag{4.2.8}$$

Finally, if $k\geq 3$ then

$$\begin{aligned}\mathrm{E}[(T(F_n)-T(F))^2]=&\frac{1}{n}\int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)\prod_{i=1}^{2} J(F(x_i))dx_i\\
&+\frac{1}{n^2}\int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)(1-2F(x_2)))\\
&\qquad\qquad J(F(x_1))J^{(1)}(F(x_2))dx_1dx_2\\
&+\frac{1}{2n^2}\int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)^2\prod_{i=1}^{2} J^{(1)}(F(x_i))dx_i\\
&+\frac{1}{4n^2}\left(\int F(x)(1-F(x))J^{(1)}(F(x))dx\right)^2\\
&+\frac{1}{n^2}\int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)F(x_2)(1-F(x_2))\\
&\qquad\qquad J(F(x_1))J^{(2)}(F(x_2))dx_1dx_2\\
&+o(1/n^2).\end{aligned} \tag{4.2.9}$$

The remainder of this section consists of the proofs of propositions 4.2.5 and 4.2.6.

We will need the following version of Taylor's theorem which can be found in Hardy (1952) to prove Fréchet differentiability for both L- and M-estimates.

Lemma 4.2.7. *If f has a finite k^{th} derivative at x then*

$$f(x+t) = \sum_{j=0}^{k} f^{(j)}(x)t^j/j! + o(|\, t\,|^k).$$

Proof of proposition 4.2.5. To obtain differentials we recall (4.2.4) and definitions 3.3.1 and 3.3.2. If $G_i \in \mathcal{D}$, $i = 1, 2, \ldots, j$, $1 \le j \le k$, then the following interchange of integration is justified under the assumptions of the proposition:

$$\begin{aligned} T_j(F; G_1, \ldots, G_j) &= \int \cdots \int \left(-\int_{\max x_i}^{\infty} J^{(j-1)}(F(y))dy \right) \prod_{i=1}^{j} dG_i(x_i) \\ &= \int \cdots \int \left(-\int \left(\prod_{i=1}^{j} \delta_{x_i}(y) \right) J^{(j-1)}(F(y))dy \right) \prod_{i=1}^{j} dG_i(x_i) \\ &= -\int \left(\prod_{i=1}^{j} \int \delta_{x_i}(y) dG_i(x_i) \right) J^{(j-1)}(F(y))dy \\ &= -\int \left(\prod_{i=1}^{j} G_i(y) \right) J^{(j-1)}(F(y))dy. \end{aligned} \tag{4.2.10}$$

Let $0 < \epsilon < \min(\delta_1, \delta_2)$. There exist a and b such that

$$-\infty < a < F^{-1}(\delta_1 - \epsilon) < F^{-1}(1 - \delta_2 + \epsilon) < b < \infty.$$

If $\|\, G \,\|_\infty < \epsilon$ then for $x \le a$ and $x \ge b$ we have

$$\int_{F(x)}^{F(x)+G(x)} J(u)du = 0.$$

For the remainder of the proof we assume that $\|\, G \,\|_\infty < \epsilon$. We now show that $\int_{F(x)}^{F(x)+G(x)} J(u)du$ has bounded total variation. For $n < \infty$, $a \le x_0 \le \cdots \le x_n \le b$

$$\begin{aligned} &\sum_{i=1}^{n} \left| \int_{F(x_i)}^{F(x_i)+G(x_i)} J(u)du - \int_{F(x_{i-1})}^{F(x_{i-1})+G(x_{i-1})} J(u)du \right| \\ &= \sum_{i=1}^{n} \left| \int_{F(x_{i-1})+G(x_{i-1})}^{F(x_i)+G(x_i)} J(u)du - \int_{F(x_{i-1})}^{F(x_i)} J(u)du \right| \\ &\le \|\, J \,\|_\infty \sum_{i=1}^{n} (2 \,|\, F(x_i) - F(x_{i-1}) \,|\, + \,|\, G(x_i) - G(x_{i-1}) \,|). \end{aligned} \tag{4.2.11}$$

Since F, $G \in D$ they are of bounded variation. It follows from (4.2.11) that $\int_{F(x)}^{F(x)+G(x)} J(u)du$ has bounded variation. We may now rewrite $T(F+G)-T(F)$ as

$$\begin{aligned} T(F+G)-T(F) &= \int xd\left(\int_0^{F(x)+G(x)} J(u)du\right) - \int xd\left(\int_0^{F(x)} J(u)du\right) \\ &= \int_{-\infty}^{\infty} xd\left(\int_{F(x)}^{F(x)+G(x)} J(u)du\right) \\ &= \int_a^b xd\left(\int_{F(x)}^{F(x)+G(x)} J(u)du\right). \end{aligned} \tag{4.2.12}$$

Applying lemma 4.2.2 we obtain

$$T(F+G)-T(F) = -\int_a^b \int_{F(x)}^{F(x)+G(x)} J(u)dudx. \tag{4.2.13}$$

For $0 \le i \le k$ equations (4.2.10) and (4.2.12) imply that the remainder of the i^{th} order Taylor approximation is

$$\begin{aligned} \mid R_i(F;G) \mid &= \left| \int_a^b \left(\int_{F(x)+G(x)}^{F(x)} J(u)du + \sum_{j=1}^{i} G^j(x) J^{(j-1)}(F(x))/j! \right) dx \right| \\ &\le \int_a^b \left| \frac{\int_{F(x)+G(x)}^{F(x)} J(u)du + \sum_{j=1}^{i} G^j(x) J^{(j-1)}(F(x))/j!}{G^i(x)} \right| dx \parallel G \parallel_\infty^i \end{aligned}$$

where the integrand of the last line is defined to be 0 when $G(x) = 0$. Thus to show that T is k times Fréchet differentiable at F it suffices to show that, for $0 \le i \le k$, as $\parallel G \parallel_\infty$ goes to zero

$$\int_a^b W_{G,i}(x)dx \equiv \int_a^b \left| \frac{\int_{F(x)+G(x)}^{F(x)} J(u)du + \sum_{j=1}^{i} G^j(x) J^{(j-1)}(F(x))/j!}{G^i(x)} \right| dx \to 0. \tag{4.2.14}$$

Let i be an integer with $0 \le i \le k$. Let $A = \{x : J^{(j)}(u)$ is continuous at $u = F(x)$, $0 \le j < k\}$. If $x \in A$ then $W_{G,i}(x) \to 0$ as $\parallel G \parallel_\infty \to 0$ by the version of Taylor's theorem given in lemma 4.2.7. From the assumption that $J^{(k-1)}$ is continuous $a.e$ with respect to F^{-1} it follows that the complement of A has Lebesgue measure zero. Under the assumptions, we have by Taylor's theorem that

$$\left| \int_{F(x)+G(x)}^{F(x)} J(u)du + \sum_{j=1}^{i-1} G^j(x) J^{(j-1)}(F(x))/j! \right| \le \parallel J^{(i-1)} \parallel_\infty G^i(x)/i!.$$

Thus $W_{G,i}$ can be bounded by $2 \parallel J^{(i-1)} \parallel_\infty$. It follows that from the dominated convergence theorem that for any sequence G_n such that $\parallel G_n \parallel_\infty$ goes to zero as $n \to \infty$ the integral $\int W_{G_n,i}(x)dx \to 0$. By a standard theorem from analysis it follows that $\int W_{G,i}(x)dx \to 0$ as $\parallel G \parallel_\infty \to 0$. ∎

Note that if $J^{(k-1)}$ is discontinuous at $u_1 \in (0,1)$ then for any $\epsilon > 0$ there exists a cdf H with $\| H - F \|_\infty < \epsilon$ and $H^{-1}(u)$ not continuous at u_1. Thus the set A of the above proof is, in this case, not of measure zero and we may not apply the dominated convergence theorem as above. This means that we cannot extend this proof directly to show Fréchet differentiability under the standard definition when $J^{(k-1)}$ is not continuous.

Proof of proposition 4.2.6. Let $X_1, X_2 \ldots$ be iid observations from F. Let F_n be the empirical cdf of the first n observations and let $X_{i:n}$ denote the i^{th} order statistic of the first n observations, $1 \leq i \leq n$. We may write

$$T(F_n) = \sum_{i=[\![\delta_1 n]\!]}^{[\![(1-\delta_2)n]\!]+1} X_{i:n} \int_{(i-1)/n}^{i/n} J(u)du.$$

This implies

$$| T(F_n) | \leq \left(| X_{[\![\delta_1 n]\!]:n} | + | X_{[\![(1-\delta_2)n]\!]+1:n} | \right) \| J \|_\infty .$$

Thus (4.2.5) follows from proposition 4.2.4 and theorem 3.4.1.

We now wish to apply theorem 3.5.3 to show the remaining results. Equation (4.2.6) follows from (3.5.8). Given the restrictions of this proposition it is easy to justify all changes of order of integration in the following calculations. Derivation of the first formula is fairly detailed; steps are left out of subsequent calculations. In each case, we begin by applying (4.2.10). Recall $Z_i = \delta_{X_i} - F$.

$$\begin{aligned}
\mathrm{E}[(T_1(F; Z_1))^2] &= \int \left(\int (\delta_y(x) - F(x)) J(F(x)) dx \right)^2 dF(y) \\
&= \int \int \int (\delta_y(x_1) - F(x_1))(\delta_y(x_2) - F(x_2)) dF(y) \prod_{i=1}^{2} J(F(x_i)) dx_i \\
&= \int \int \int \Big(\delta_y(x_1)\delta_y(x_2) - \delta_y(x_1)F(x_2) - \delta_y(x_2)F(x_1) + F(x_1)F(x_2) \Big) \\
&\qquad\qquad dF(y) \prod_{i=1}^{2} J(F(x_i)) dx_i \\
&= \int \int \Big(F(\min(x_1, x_2)) - F(x_1)F(x_2) \Big) \prod_{i=1}^{2} J(F(x_i)) dx_i.
\end{aligned}$$

Equation (4.2.7) now follows under the given conditions. This is the usual first order variance approximation for L-estimates. We continue with the bias approximation.

$$\begin{aligned}
\mathrm{E}[T_2(F; Z_1, Z_1)] &= \int \int (\delta_y(x) - F(x))^2 J^{(1)}(F(x)) dF(y) dx \\
&= \int F(x)(1 - F(x)) J^{(1)}(F(x)) dx.
\end{aligned}$$

Equation (4.2.8) now follows under the given conditions. Finally we do the remaining calculations needed for the second order mean squared error approximation.

$$\begin{aligned}
\mathrm{E}[T_1(F;Z_1)T_2(F;Z_1,Z_1)] &= \int\int\int(\delta_y(x_1)-F(x_1))(\delta_y(x_2)-F(x_2))^2 \\
&\qquad J(F(x_1))J^{(1)}(F(x_2))dx_1dx_2dF(y) \\
&= \int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)(1-2F(x_2)) \\
&\qquad J(F(x_1))J^{(1)}(F(x_2))dx_1dx_2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[(T_2(F;Z_1,Z_2))^2] &= \int\int\int\int\prod_{j=1}^{2}\left(\prod_{i=1}^{2}(\delta_{y_j}(x_i)-F(x_i))J^{(1)}(F(x_i))dx_i\right)dF(y_j) \\
&= \int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)^2\prod_{i=1}^{2}J^{(1)}(F(x_i))dx_i
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[T_1(F;Z_1)T_3(F;Z_1,Z_2,Z_2)] &= \int\int\Big(F(\min(x_1,x_2))-F(x_1)F(x_2)\Big)F(x_2)(1-F(x_2)) \\
&\qquad J(F(x_1))J^{(2)}(F(x_2))dx_1dx_2.
\end{aligned}$$

Equation (4.2.9) now follows under the given conditions. ∎

§4.3 Theory for M-estimates.

In this section we will show the Fréchet differentiability of many M-estimates and give the theory for first and second order variance approximations. Boos and Serfling (1980) have given conditions under which an M-functional is one time Fréchet differentiable. We give a somewhat different proof for conditions under which an M-functional is one, two, or three times Fréchet differentiable.

For notational convenience we shall write $\int\psi dF$ instead of $\int_{-\infty}^{\infty}\psi(x)dF(x)$. In general when the argument of a function in an integrand is left out it will simply be the running variable (in this case x) and when the limits of integration are not given they will be $\pm\infty$. Otherwise arguments of functions and limits of integration will be explicitly given.

Definition 4.3.1. *Let ψ be a real-valued function on $\Re$. An* **M-estimate** *corresponding to ψ is $T(F_n)$ where $T(F)$ is any functional which satisfies*

$$0=\int\psi(x-T(F))dF(x)$$

on its domain of definition.

We now give a heuristic method of finding differentials for M-estimates. The usual candidate for the j^{th} differential is

$$T_j(F;G) = \frac{d^j}{dh^j} T(F+hG)\,|_{h=0}\,. \tag{4.3.1}$$

For any $G \in \mathcal{D}$ we define

$$\lambda_G(t) = \int \psi(x-t)dG(x).$$

Let $\mathcal{F}$ denote the domain of definition of T. For $G \in \mathcal{F}$ we have $\lambda_G(T(G)) = 0$. Noting that

$$\lambda_{F+hG}(t) = \lambda_F(t) + h\lambda_G(t)$$

we have thus

$$0 = \lambda_F(T(F+hG)) + h\lambda_G(T(F+hG)).$$

Differentiating this with respect to h we obtain

$$0 = \lambda_F'(T(F+hG))\frac{d}{dh}T(F+hG) + \lambda_G(T(F+hG)) + h\lambda_G'(T(F+hG))\frac{d}{dh}T(F+hG). \tag{4.3.2}$$

Setting $h=0$, this and (4.3.1) yield

$$T_1(F;G) = -\frac{\lambda_G(T(F))}{\lambda_F'(T(F))}. \tag{4.3.3}$$

Differentiating both sides of (4.3.2) with respect to h again we obtain

$$\begin{aligned} 0 =& \lambda_F''(T(F+hG))\left(\frac{d}{dh}T(F+hG)\right)^2 + \lambda_F'(T(F+hG))\frac{d^2}{dh^2}T(F+hG) \\ &+ 2\lambda_G'(T(F+hG))\frac{d}{dh}T(F+hG) + h\lambda_G''(T(F+hG))\left(\frac{d}{dh}T(F+hG)\right)^2 \\ &+ h\lambda_G'(T(F+hG))\frac{d^2}{dh^2}T(F+hG). \end{aligned}$$

Setting $h=0$ in this equation and applying (4.3.1) and (4.3.3) we obtain a candidate for the second differential:

$$\begin{aligned} T_2(F;G) &= -\frac{\lambda_F''(T(F))(T_1(F;G))^2 + 2\lambda_G'(T(F))T_1(F;G)}{\lambda_F'(T(F))} \\ &= 2\frac{\lambda_G(T(F))\lambda_G'(T(F))}{(\lambda_F'(T(F)))^2} - \frac{(\lambda_G(T(F)))^2\lambda_F''(T(F))}{(\lambda_F'(T(F)))^3}. \end{aligned} \tag{4.3.4}$$

The candidate for the third differential is obtained in the same fashion:

$$
\begin{aligned}
T_3(F;G) = &-\Big(\lambda_F'''(T(F))(T_1(F;G))^3 + 3\lambda_F''(T(F))T_1(F;G)T_2(F;G) \\
&+ 3\lambda_G''(T(F))(T_1(F;G))^2 + 3\lambda_G'(T(F))T_2(F;G)\Big)(\lambda_F'(T(F)))^{-1} \\
= &\frac{(\lambda_G(T(F)))^3\lambda_F'''(T(F))}{(\lambda_F'(T(F)))^4} + 9\frac{(\lambda_G(T(F)))^2\lambda_G'(T(F))\lambda_F''(T(F))}{(\lambda_F'(T(F)))^4} \\
&- 3\frac{(\lambda_G(T(F)))^3(\lambda_F''(T(F)))^2}{(\lambda_F'(T(F)))^5} - 3\frac{(\lambda_G(T(F)))^2\lambda_G''(T(F))}{(\lambda_F'(T(F)))^3} - 6\frac{\lambda_G(T(F))(\lambda_G'(T(F)))^2}{(\lambda_F'(T(F)))^3}.
\end{aligned}
\tag{4.3.5}
$$

We will give an example of how to obtain kernels from (4.3.3)—(4.3.5). We consider T_2. A bilinear function $T_2(F;G_1,G_2)$ satisfying $T_2(F;G) = T_2(F;G,G)$ where $T_2(F;G)$ is as in (4.3.4) is

$$
T_2(F;G_1,G_2) = \frac{\lambda_{G_1}(T(F))\lambda_{G_2}'(T(F)) + \lambda_{G_2}(T(F))\lambda_{G_1}'(T(F))}{(\lambda_F'(T(F)))^2} - \frac{\lambda_{G_1}(T(F))\lambda_{G_2}(T(F))\lambda_F''(T(F))}{(\lambda_F'(T(F)))^3}.
$$

Switching the order of integration and differentiation, and replacing G_1 and G_2 with distributions degenerate at x_1 and x_2 respectively, we obtain a candidate for the second kernel (assuming $T(F) = 0$)

$$
h_2(F;x_1,x_2) = -\frac{\psi(x_1)\psi'(x_2) + \psi(x_2)\psi'(x_1)}{(\int \psi' dF)^2} + \frac{\psi(x_1)\psi(x_2)\int \psi'' dF}{(\int \psi' dF)^3}.
$$

We will now give sufficient conditions for a functional corresponding to an M-estimate to be one, two or three times Fréchet differentiable according to definition 3.3.2 (recall this has slightly weaker requirements than the standard definition). No additional concepts should be needed to extend the proof of the theorem to higher order differentials.

Proposition 4.3.2. *Let F be a cdf. Let ψ be such that $\int \psi dF = 0$. Let k be 1, 2, or 3. Assume that ψ and its first $k-1$ derivatives are absolutely continuous everywhere and have bounded total variation. Assume that the k^{th} derivative of ψ exists a.e. with respect to Lebesgue measure and with respect to the measure corresponding to F, and has bounded total variation. This implies that the integral $\int \psi' dF$ is well defined. Assume $\int \psi' dF > 0$. For $G \in \mathcal{D}$ with $\| G \|_\infty$ sufficiently small there exists a functional $T(F+G)$ such that $0 = \int \psi(x - T(F+G))d(F(x) + G(x))$. Let $\mathcal{F}$ be a neighborhood of F in $\mathcal{D}$ where such a T can be defined. Then T is k times Fréchet differentiable at F with respect to $\| \cdot \|_\infty$ and (4.3.3)—(4.3.5) hold. The kernels of the first three differentials of T are*

$$
h_1(F;x) = \frac{\psi(x)}{\int \psi' dF}, \tag{4.3.6}
$$

$$h_2(F;x_1,x_2) = -\frac{\psi(x_1)\psi'(x_2)+\psi'(x_1)\psi(x_2)}{(\int \psi' dF)^2} + \psi(x_1)\psi(x_2)\frac{\int \psi'' dF}{(\int \psi' dF)^3}, \tag{4.3.7}$$

and

$$\begin{aligned} h_3(F;x_1,x_2,x_3) = &-\psi(x_1)\psi(x_2)\psi(x_3)\frac{\int \psi''' dF}{(\int \psi' dF)^4} \\ &-3\Big(\psi(x_1)\psi(x_2)\psi'(x_3)+\psi(x_1)\psi'(x_2)\psi(x_3)+\psi'(x_1)\psi(x_2)\psi(x_3)\Big)\frac{\int \psi'' dF}{(\int \psi' dF)^4} \\ &+3\psi(x_1)\psi(x_2)\psi(x_3)\frac{(\int \psi'' dF)^2}{(\int \psi' dF)^5} \\ &+\frac{\psi(x_1)\psi(x_2)\psi''(x_3)+\psi(x_1)\psi''(x_2)\psi(x_3)+\psi''(x_1)\psi(x_2)\psi(x_3)}{(\int \psi' dF)^3} \\ &+2\frac{\psi(x_1)\psi'(x_2)\psi'(x_3)+\psi'(x_1)\psi(x_2)\psi'(x_3)+\psi'(x_1)\psi'(x_2)\psi(x_3)}{(\int \psi' dF)^3}. \end{aligned} \tag{4.3.8}$$

We can apply this proposition to obtain a result on moment approximations of M-estimates with non-decreasing ψ functions. We restrict ourselves to the case of the underlying distribution being symmetric. This is not necessary, but is a common assumption and makes the expressions simpler.

Proposition 4.3.3. *Let F be a cdf which is symmetric about 0 such that*

$$0 < \alpha = \liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log x}.$$

Let ψ be a bounded, non-decreasing, absolutely continuous, odd function. Assume that ψ' exists and is continuous a.e. with respect to Lebesgue measure and with respect to the measure corresponding to F, is of bounded total variation, and satisfies $\int \psi' dF > 0$. Let $\mathcal{F} \in \mathcal{D}$ be the set where a solution $T(G)$ of $0 = \int \psi(x - T(G))dG(x)$ exists. Then $\mathcal{F}$ contains all $G \in \mathcal{D}$ which are non-decreasing and a neighborhood of F with respect to $\|\cdot\|_\infty$. If G is non-negative and non-decreasing this solution minimizes $\int \rho(x-T(G))dG(x)$ where $\rho(x) = \int_0^x \psi(y)dy$. Letting F_n denote the empirical cdf of n iid observations from F it follows that $T(F_n)$ minimizes $\int \rho(x - T(F_n))dF_n$ and

$$\mathrm{E}[T(F_n)] = 0, \quad \mathrm{E}[(T(F_n))^2] = \frac{1}{n}\frac{\int \psi^2 dF}{(\int \psi' dF)^2} + o(1/n). \tag{4.3.9}$$

Assume, in addition, that the first two derivatives of ψ are absolutely continuous and bounded. Assume also that the third derivative of ψ exists, is continuous a.e. with respect to Lebesgue measure and with respect to the measure corresponding to F, and has bounded total variation. Then

$$\begin{aligned} \mathrm{E}[(T(F_n))^2] = &\frac{n-1}{n^2}\frac{\int \psi^2 dF}{(\int \psi' dF)^2} - \frac{2}{n^2}\frac{\int \psi^2\psi' dF}{(\int \psi' dF)^3} + \frac{3}{n^2}\frac{\int \psi^2 dF \int (\psi')^2 dF}{(\int \psi' dF)^4} \\ &+\frac{3}{n^2}\frac{\int \psi^2 dF \int \psi\psi'' dF}{(\int \psi' dF)^4} - \frac{1}{n^2}\frac{(\int \psi^2 dF)^2 \int \psi''' dF}{(\int \psi' dF)^5} + o(1/n^2). \end{aligned} \tag{4.3.10}$$

The remainder of this section is devoted to proving the above two propositions. We begin by giving a series of three lemmas.

Lemma 4.3.4. *Suppose ψ is a function of bounded total variation. Then there exists $A > 0$ such that for all $G \in \mathcal{D}$*

$$\sup_{-\infty<t<\infty} \mid \int \psi(x-t)dG(x) \mid < A \parallel G \parallel_\infty .$$

Proof: Since ψ has bounded total variation it may be written as the difference of two bounded non-decreasing functions ψ_1 and ψ_2. Let A be such that $A/2$ is a bound for these two functions. For any fixed value of t

$$\mid \int \psi(x-t)dG(x) \mid \leq \mid \int \psi_1(x-t)dG(x) \mid + \mid \int \psi_2(x-t)dG(x) \mid .$$

But for $i = 1, 2$

$$\mid \int \psi_i(x-t)dG(x) \mid \leq (A/2) \sup_{-\infty<y<\infty} \mid \int_{-\infty}^{y} dG(x) \mid \leq (A/2) \parallel G \parallel_\infty$$

and the contention follows since t was arbitrary. ∎

Lemma 4.3.5. *Let $F \in \mathcal{D}$. Suppose ψ is absolutely continuous and bounded, and that ψ' exists a.e. F and Lebesgue, and is bounded where it exists. Then*

$$\frac{d}{dt} \int \psi(x-t)dF(x) \Big|_{t=0} = - \int \psi'(x)dF(x).$$

If, in addition, ψ is differentiable at each real x then for any $G \in \mathcal{D}$, $t \in \Re$

$$\frac{d}{dt} \int \psi(x-t)dG(x) = - \int \psi'(x-t)dG(x).$$

Proof: By the definition of derivatives

$$\frac{d}{dt} \int \psi(x-t)dF(x) = \lim_{h\to 0} \int \frac{\psi(x-t-h) - \psi(x-t)}{h} dF(x).$$

For $t = 0$ the limit of the integrand of the right hand side is $-\psi'(x)$ except at a set of F measure zero. Since the integrand is bounded by $\sup_{-\infty<y<\infty} \mid \psi'(y) \mid$ the first result follows by the dominated convergence theorem.

The proof is the same for the second claim since $\psi'(x-t)$ exists for *a.e.* $x \in G$ for any $G \in \mathcal{D}$ under the additional assumption. ∎

Our proof of Fréchet differentiability uses this lemma. If ψ is not continuously differentiable then for any $\epsilon > 0$ there exists $H \in D$ with $\| H - F \|_\infty < \epsilon$ such that H is not continuous at a point of discontinuity of ψ'. Thus we cannot use this lemma to show the standard definition of Fréchet differentiability holds if ψ is not continuously differentiable.

Lemma 4.3.6. *Let F be a cdf. Suppose ψ is an absolutely continuous function with bounded total variation whose derivative exists, is bounded, and is continuous a.e. F. Suppose*

$$\lambda_F(t) = \int \psi(x-t)dF(x),$$

$\lambda_F(0) = 0$, and $\lambda'_F(0) < 0$. Then there exists a functional $T(G)$ in a neighborhood of F in D such that

$$\lambda_G(T(G)) = \int \psi(x - T(G))dG(x) = 0$$

and

$$| T(G) - T(F) | = O(\| G - F \|_\infty).$$

Proof: From lemma 4.3.4 there exists $A > 0$ such that

$$\begin{aligned}\sup_{-\infty<t<\infty} | \lambda_G(t) - \lambda_F(t) | &= \sup_{-\infty<t<\infty} | \textstyle\int \psi(x-t)(dG(x) - dF(x)) | \\ &\le A \| G - F \|_\infty .\end{aligned}$$

Let $\delta_1 > 0$, $0 < \delta_2 < -\lambda'_F(0)$ be such that if $t \in (-\delta_1, \delta_1)$ and $t \neq 0$ then $\lambda_F(t)/t < \lambda'_F(0) + \delta_2$. For $t \in (0, \delta_1)$

$$\begin{aligned}\lambda_G(t) &\le \lambda_F(t) + A \| G - F \|_\infty \\ &\le t(\lambda'_F(0) + \delta_2) + A \| G - F \|_\infty .\end{aligned}$$

It follows that if $-A \| G - F \|_\infty /(\lambda'_F(0) + \delta_2) < t < \delta_1$ then $\lambda_G(t) < 0$. Similarly, if $A \| G - F \|_\infty /(\lambda'_F(0) + \delta_2) > t > -\delta_1$ then $\lambda_G(t) > 0$. Since $\lambda_G(t)$ is continuous in t for each $G \in D$, there is a solution of $\lambda_G(t) = 0$ with $| t | < -A \| G - F \|_\infty /(\lambda'_F(0) + \delta_2)$ if $\| G - F \|_\infty < -\delta_1(\lambda'_F(0) + \delta_2)/A$. ∎

Proof of proposition 4.3.2. Under the assumptions, we have by lemma 4.3.5 that, for $j = 1, \ldots, k$,

$$\left.\frac{d^j}{dt^j}\lambda_F(t)\right|_{t=0} = \left.\frac{d^j}{dt^j}\int \psi(x-t)dF(x)\right|_{t=0} = (-1)^j \int \psi^{(j)}dF.$$

We also have for arbitrary $G \in D$, $t \in \Re$, and $j = 1, \ldots, k-1$,

$$\frac{d^j}{dt^j}\lambda_G(t) = \frac{d^j}{dt^j}\int \psi(x-t)dG(x) = (-1)^j \int \psi^{(j)}(x-t)dG.$$

Using these facts it is easy to show that the formulas (4.3.3)—(4.3.5) hold.

For the remainder of the proof assume that $\| G - F \|_\infty$ is sufficiently small so that $T(G)$ is well defined. By assumption $\lambda'_F(0) = -\int \psi' dF > 0$. Thus we may show

$$-\lambda'_F(0)R_k(F; G-F) = -\lambda'_F(0)\left(T(G) - T(F) - \sum_{j=1}^{i} T_j(F; G-F)\right)$$

is $o(\| G - F \|_\infty^i)$, $0 \le i \le k$ to show Fréchet differentiability. For $i = 0$ the result follows from lemma 4.3.6. Recall $\lambda_F(0) = 0$, $\lambda_G(T(G)) = 0$. From (4.3.3) we have

$$\begin{aligned} -\lambda'_F(0)R_1(F; G-F) &= -\lambda'_F(0)T(G) - \lambda_G(0) \\ &= \lambda_F(T(G)) - \lambda'_F(0)T(G) \\ &\quad + \lambda_G(T(G)) + \lambda_F(0) - \lambda_F(T(G)) - \lambda_G(0). \end{aligned} \tag{4.3.11}$$

By lemma 4.3.6, $T(G) = T(G) - T(F) = O(\| G - F \|_\infty)$ and thus $\lambda_F(T(G)) - \lambda'_F(0)T(G) = o(\| G - F \|_\infty)$. We may rewrite the last line of (4.3.11) as

$$\int \Big(\psi(x - T(G)) - \psi(x)\Big) d(G - F). \tag{4.3.12}$$

Since ψ is uniformly continuous and $T(G) = O(\| G - F \|_\infty)$ this may be written as

$$O(\| G - F \|_\infty) \int d \mid G - F \mid = O(\| G - F \|_\infty^2).$$

Applying the triangle inequality it follows that T is one time Fréchet differentiable.

We continue with similar proofs for second and third differentials. First we note that for $k > 1$

$$\begin{aligned} -\lambda'_F(0)R_k(F; G-F) &= -\lambda'_F(0)\left(T(G) - T(F) - \sum_{j=1}^{k} T_j(F; G-F)/j!\right) \\ &= -\lambda'_F(0)R_{k-1}(F; G-F) + \lambda'_F(0)T_k(F; G-F)/k!. \end{aligned} \tag{4.3.13}$$

From (4.3.4), (4.3.11) and (4.3.13) we have (after rearrangement)

$$\begin{aligned} -\lambda'_F(0)R_2(F; G-F) =& \lambda_F(T(G)) - \sum_{j=1}^{2} \frac{(T(G))^2}{j!} \lambda_F^{(j)}(0) \\ &- \Big(\lambda_G(0) + \lambda_F(T(G)) + (\lambda'_G(0) - \lambda'_F(0))T(G)\Big) \\ &+ \frac{1}{2}\lambda''_F(0)\Big((T(G))^2 - (T_1(F; G-F))^2\Big) \\ &+ \Big(\lambda'_G(0) - \lambda'_F(0)\Big)\Big(T(G) - T_1(F; G-F)\Big). \end{aligned} \tag{4.3.14}$$

The first line of this is a second order Taylor's series expansion and is $o(\| G-F \|_\infty^2)$ by the version of Taylor's theorem given in lemma 4.2.7.

The second line of (4.3.14) may be rewritten as

$$\int\Big(\psi(x-T(G))-\psi(x)+\psi'(x)T(G)\Big)d(G-F). \tag{4.3.15}$$

Under the condition that ψ' is continuous everywhere and of bounded total variation, the integrand may be bounded uniformly by a constant times $(T(G))^2$. Thus (4.3.15) can be shown to be $O(\| G-F \|_\infty^3)$ as (4.3.12) was shown to be $O(\| G-F \|_\infty^2)$.

To show the third line of (4.3.14) is $o(\| G-F \|_\infty^2)$ we note that we have shown

$$T(G)=T_1(F;G-F)+o(\| G-F \|_\infty)$$

and thus

$$(T(G))^2=(T_1(F;G-F))^2+o(\| G-F \|_\infty)T_1(F;G-F)+o(\| G-F \|_\infty^2).$$

But $T_1(F;G-F)=O(\| G-F \|_\infty)$ by lemma 3.4.6.

In the fourth line of (4.3.14) we have

$$\lambda'_G(0)-\lambda'_F(0)=-\int\psi' d(G-F)=O(\| G-F \|_\infty)$$

and $T(G)-T_1(F;G-F)=o(\| G-F \|_\infty)$ which implies the product is $o(\| G-F \|_\infty^2)$.

Applying the triangle inequality we see that $-\lambda'_F(0)R_2(F;G-F)=o(\| G-F \|_\infty^2)$ and we have shown that T is two times Fréchet differentiable at F.

From (4.3.5), (4.3.13) and (4.3.14) we have (after rearrangement)

$$\begin{aligned}
-\lambda'_F(0)R_3(F;G-F) =&\lambda_F(T(G))-\sum_{j=1}^{3}\frac{(T(G))^j}{j!}\lambda_F^{(j)}(0)\\
&-\Big(\lambda_G(0)+\lambda_F(T(G))+\sum_{j=1}^{2}\frac{(T(G))^j}{j!}(\lambda_G^{(j)}(0)-\lambda_F^{(j)}(0))\Big)\\
&+\Big(\lambda'_G(0)-\lambda'_F(0)\Big)\Big(T(G)-T_1(F;G-F)-\frac{1}{2}T_2(F;G-F)\Big)\\
&+\frac{1}{2}\Big(\lambda''_G(0)-\lambda''_F(0)\Big)\Big((T(G))^2-(T_1(F;G-F))^2\Big)\\
&+\frac{1}{2}\lambda''_F(0)\Big((T(G))^2-(T_1(F;G-F))^2-T_1(F;G-F)T_2(F;G-F)\Big)\\
&+\frac{1}{6}\lambda'''_F(0)\Big(((T(G))^3-(T_1(F;G-F))^3\Big).
\end{aligned} \tag{4.3.16}$$

The ideas needed to show that each line of the right hand side of (4.3.16) is $o(\| G-F \|_\infty^3)$ have already been given. We omit the details. ∎

Proof of proposition 4.3.3. We have shown in proposition 4.3.2 that $T(G)$ is well defined in a neighborhood of F. Let G be a non-decreasing element of $\mathcal{D}$. As $t \to \infty$, $\lambda_G(t) = \int \psi(x-t)dG < 0$ and as $t \to -\infty$, $\lambda_G(t) > 0$. Since $\lambda_G(t)$ is continuous and non-increasing this implies that there is a solution $T(G)$ of the equation $\lambda_G(t) = 0$. This also implies $\lambda_G(t) \geq 0$ for all $t < T(G)$ and $\lambda_G(t) \leq 0$ for all $t > T(G)$. Since ρ is convex we have

$$\int \rho(x-t)dG(x) \geq \int \rho(x - T(G))dG(x) + (T(G)-t)\int \psi(x-t)dG(x). \tag{4.3.17}$$

This holds even if $\int \rho(x - T(G))dG(x) = \infty$ since ρ is non-negative and G is totally positive. The second term of the right hand side of (4.3.17) is greater than or equal to zero, and thus $t = T(G)$ minimizes $\int \rho(x-t)dG(x)$.

We will now justify the application of theorem 3.4.1. Proposition 4.3.2 gives us the existence of the appropriate differentials. The tail condition needed on F is assumed. Let A be such that if $x \geq A$ then $\psi(x) \geq (1/2)\sup\psi(y) \equiv B/2$. Let $\epsilon = \min(1 - F(A), 1/4)$. Then

$$\int \psi(x - X_{[\![\epsilon n]\!]:n} + A)dF_n \geq \frac{3B}{8} - \frac{B}{4} = \frac{B}{8} > 0$$

which implies $T(F_n) \geq X_{[\![\epsilon n]\!]:n} - A$. Similarly $T(F_n) \leq X_{[\![(1-\epsilon)n]\!]+1:n} + A$. Thus $\mid T(F_n) \mid \leq \mid X_{[\![\epsilon n]\!]:n} \mid + \mid X_{[\![(1-\epsilon)n]\!]+1:n} \mid + 2A$ and condition *ii* of theorem 3.4.1 is satisfied.

We may now apply theorem 3.5.3. Under the symmetry assumptions we have $\int \psi'' dF = 0$ and thus the kernels are simplified as follows:

$$\begin{aligned}
h_1(F;x) &= \frac{\psi(x)}{\int \psi' dF} \\
h_2(F;x_1,x_2) &= -\frac{\psi(x_1)\psi'(x_2) + \psi'(x_1)\psi(x_2)}{(\int \psi' dF)^2} \\
h_3(F;x_1,x_2,x_3) &= -\psi(x_1)\psi(x_2)\psi(x_3)\frac{\int \psi''' dF}{(\int \psi' dF)^4} \\
&\quad + \frac{\psi(x_1)\psi(x_2)\psi''(x_3) + \psi(x_1)\psi''(x_2)\psi(x_3) + \psi''(x_1)\psi(x_2)\psi(x_3)}{(\int \psi' dF)^3} \\
&\quad + 2\frac{\psi(x_1)\psi'(x_2)\psi'(x_3) + \psi'(x_1)\psi(x_2)\psi'(x_3) + \psi'(x_1)\psi'(x_2)\psi(x_3)}{(\int \psi' dF)^3}.
\end{aligned}$$

From definition 3.3.1 and theorem 3.5.3 the calculations to show the desired results are now simple. We give one example. We have (recall $Z_i = \delta_{X_i} - F$)

$$T_1(F;Z_1) = \frac{\psi(X_1) - \int \psi dF}{\int \psi' dF} = \frac{\psi(X_1)}{\int \psi' dF},$$

$$T_2(F;Z_1,Z_1) = -2\frac{\psi(X_1)\psi'(X_1) - \psi(X_1)\int \psi' dF - \psi'(X_1)\int \psi dF + \int \psi dF \int \psi' dF}{\left(\int \psi' dF\right)^2}.$$

Thus

$$\mathrm{E}[T_1(F;Z_1)T_2(F;Z_1,Z_1)] = -2\frac{\int \psi^2\psi' dF}{\left(\int \psi' dF\right)^3} + 2\frac{\int \psi^2 dF}{\left(\int \psi' dF\right)^2}.$$

∎

Chapter 5

Applications

§5.1 Introduction.

We are at last prepared to try to give the reader a feel for the applications and limitations of the theory we have presented by discussing some numerical examples and counterexamples. We do not make any sort of extensive study of estimators or seriously attempt to find any optimal estimators. Work in this direction is being done by Eynon (1982).

In section 2 we present some valid applications of our theory. We demonstrate using Monte Carlo studies that the second order variance approximations yield big improvements over the first order approximations in some cases. Often the second order variance approximation is 'better' than Monte Carlo approximation because the amount of calculation needed to obtain the same degree of accuracy by simulation is large.

In section 3, we give some initial simulation results for nonparametric variance estimates which are first and second order expansions for variances evaluated at the empirical cdf.

In section 4 we consider the median and trimmed means. These statistics do not satisfy the Fréchet differentiability conditions of theorem 3.4.1. We show that we do not necessarily obtain valid variance expansions in these cases by taking the limit of expansions for statistics which do meet the differentiability conditions and approach the desired statistic.

§5.2 Initial examples.

In this section we give numerical examples of approximations which apply the theory of chapter 4 for small to moderate sample sizes. Using simulation we show that in the cases presented

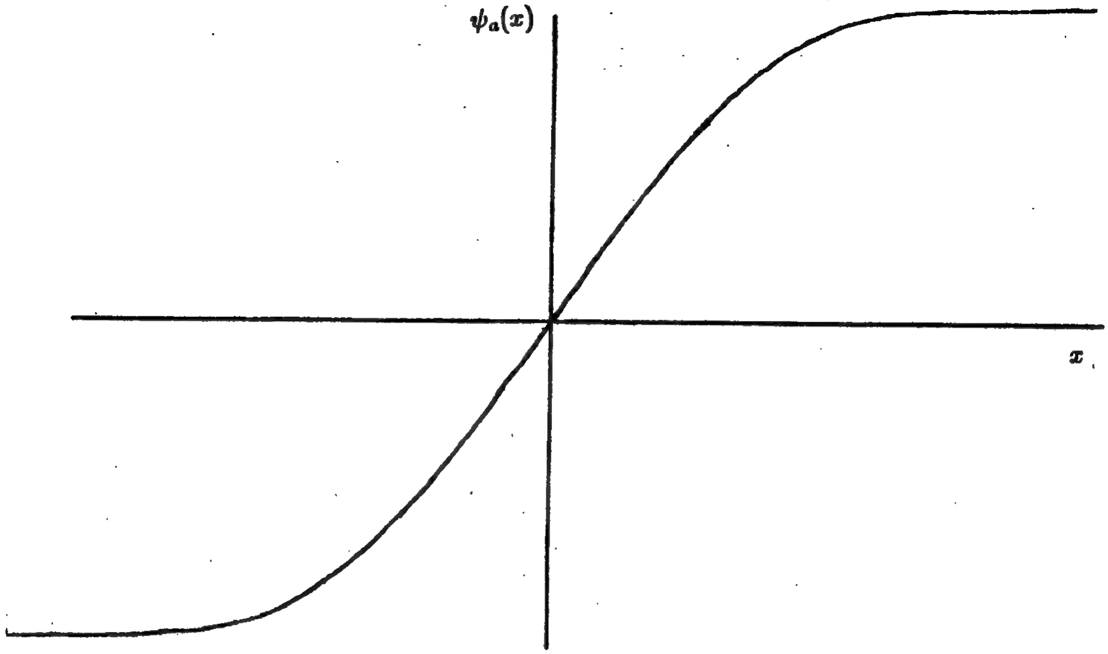

Figure 1. Graph of ψ_a as defined in (5.2.1).

these approximations are quite good. Various applications are suggested. We use the incomplete beta function

$$I_u(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^u x^{p-1}(1-x)^{q-1}dx$$

to define our estimators. The functions defining our M-estimates (recall definition 4.3.1) are defined for positive a by

$$\psi_a(x) = \begin{cases} I_{(1+x/a)/2}(3,3) - 1/2, & \text{if } |x| \leq a, \\ 1/2, & \text{if } x \geq a, \\ -1/2, & \text{if } x \leq -a. \end{cases} \tag{5.2.1}$$

Figure 1 presents a graph of ψ_a.

The first three derivatives of ψ_a are piecewise polynomial. The first two derivatives of ψ_a are continuous everywhere, and the third derivative does not exist at $\pm a$ and is continuous elsewhere.

The three distribution functions we will use in our examples are the standard normal,

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy, \tag{5.2.2}$$

the Cauchy

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x), \tag{5.2.3}$$

and the Laplace

$$F(x) = \begin{cases} 1 - \frac{1}{2}e^{-x}, & \text{if } x \geq 0, \\ \frac{1}{2}e^{x}, & \text{otherwise.} \end{cases} \tag{5.2.4}$$

All of the conditions of proposition 4.3.3 needed to obtain first and second order variance approximations are easily verified for any positive a and any of the above three distribution functions. Although our theory does apply to distributions with discrete components, we do not give any examples using such distributions.

For each of our examples we compare first and second order variance expansions from proposition 4.3.3 (for M-estimates) or proposition 4.2.6 (for L-estimates) with variance approximations obtained by simulation. In our simulations we use a combination of a linear congruential and Fibonacci pseudo random number generators as recommended in Knuth (1969), (see pp. 9, 26, 30) to generate uniform pseudo random numbers. To obtain normal pseudo random numbers we use the Box-Muller (1959) transformation in combination with the above uniform generator. To obtain Cauchy and Laplace pseudo random variables we divide normal pseudo random variables by independent random variables having the appropriate distributions. See Andrews, *et. al.* (1972), pp. 56-57. We use these so called normal/independent generators so that we may use the variance reduction techniques described in Andrews *et. al.* (1972), and (more thoroughly) in Simon (1976). For the L-estimates we present we use precisely these variance reduction techniques. For M-estimates we must use a slightly different procedure as our M-estimates are not scale invariant. The difference in swindling techniques may sometimes cause the standard error of simulation approximations for variances of M-estimates to be slightly larger than those for L-estimates.

Table 1 presents comparisons of approximations done by simulation and by the expansions of proposition 4.3.3. Except for the first order approximation with $n = 10$ in the Cauchy example all of the expansion approximations appear to be quite good. Although we have not given (complete) theoretical justification, we suspect that the behavior of the variance expansions can be described as

$$n\mathrm{Var}(T(F_n)) = \sigma_1^2 + (1/n)\sigma_2^2 + (1/n^2)\sigma_3^2 + o(n^{-2}).$$

Given that this formula is valid we would expect that the error in the first order approximation is approximately σ_2^2/n which is halved as n doubles, and the error in the second order approximation is approximately σ_3^2/n^2 which is quartered as n doubles. Because of the standard error of the simulation approximations it is difficult to tell if the error is behaving like this in many cases. However, the error in the first order expansion for the Cauchy and Laplace examples does appear to halve as n doubles and for the Cauchy example it appears that the error of the second order approximation is quartered as n doubles.

F, a	Variance Approximation	$n = 10$	$n = 20$	$n = 40$
Normal Distribution $a = 1$	Simulation size	40,000	20,000	10,000
	Simulation (Std. Err.)	1.197 (.002)	1.203 (.002)	1.203 (.003)
	1^{st} order	1.208	1.208	1.208
	2^{nd} order	1.200	1.204	1.206
Cauchy Distribution $a = .6$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	3.341 (.017)	2.714 (.012)	2.472 (.012)
	1^{st} order	2.278	2.278	2.278
	2^{nd} order	2.959	2.619	2.449
Laplace Distribution $a = 1.5$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	1.419 (.004)	1.335 (.005)	1.295 (.007)
	1^{st} order	1.266	1.266	1.266
	2^{nd} order	1.409	1.338	1.302

Table 1. Variance approximations for M-estimates with $\psi = \psi_a$.

The first column of table 1 gives the distribution for which estimates of location are being made as well as the parameter a of this M-estimate. The column labelled 'Variance approximation' gives a brief description of the values presented in each row. The last three columns are headed by the sample size for the estimates considered. The first row of each section presents the simulation size used for the Monte Carlo approximation of n times the variance given in the second row. The standard error of this approximation is given in parentheses. The third and fourth rows give the first and second order approximations, respectively, of n times the variance obtained from proposition 4.3.3.

For a given distribution function F and positive a we will consider an L-estimate as defined in definition 4.2.6 with

$$J_{a,F}(u) = \frac{\psi_a'(F^{-1}(u))}{\int \psi_a' dF} \tag{5.2.5}$$

where ψ_a' is the derivative of ψ_a in (5.2.1). Figure 2 presents a graph of this function when F is the standard normal distribution.

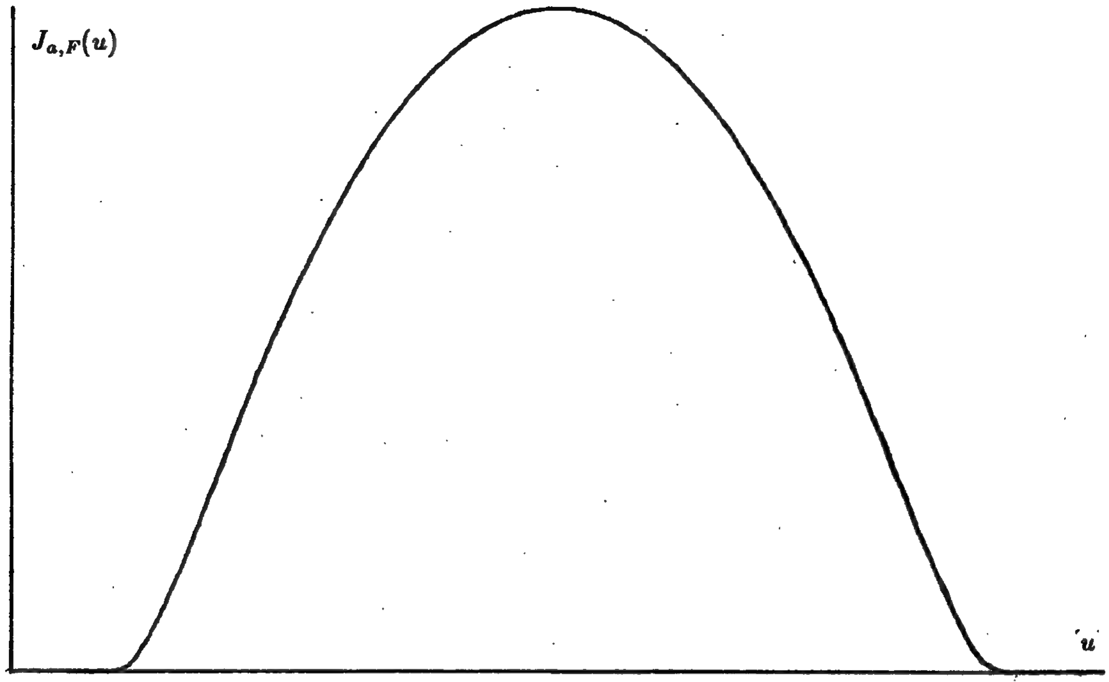

Figure 2. Graph of $J_{a,F}$ as defined in (5.2.5) with $a = 1.3$ and F =standard normal cdf.

Note that the shape of $J_{a,F}$ changes as a and F change. For F as in (5.2.2)—(5.2.4) it is easily verified that we can apply proposition 4.2.6 to obtain first and second order variance approximations of L-estimates corresponding to (5.2.5). From (4.2.4)and (4.3.6) we see that this L-estimate has the same influence curve as the M-estimate corresponding to ψ_a. This implies that the M- and L-estimates corresponding to ψ_a have the same first order variance approximation (recall (3.5.10) and the fact that $T_1(F; Z_1) \equiv h_1(F; X_1) - \mathrm{E}[h_1(F; X_1)]$, where $h_1(F; \cdot)$ is the influence curve).

Table 2 shows results similar to those given in table 1. Note that the second order term can be important for comparing variances of corresponding M- and L-estimates, especially for the Cauchy distribution. In comparing the normal examples of tables 1 and 2 we see first order asymptotically equivalent estimators where the variance for n =10, 20,40 is smaller for the L-estimate than the M-estimate. In all other examples given in this section the M-estimate has outperformed its counterpart.

F, a	Variance Approximation	$n = 10$	$n = 20$	$n = 40$
Normal Distribution $a = 1$	Simulation size	40,000	20,000	10,000
	Simulation (Std. Err.)	1.181 (.001)	1.195 (.002)	1.198 (.003)
	1^{st} order	1.208	1.208	1.208
	2^{nd} order	1.182	1.195	1.202
Cauchy Distribution $a = .6$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	3.404 (.016)	2.743 (.012)	2.484 (.012)
	1^{st} order	2.278	2.278	2.278
	2^{nd} order	3.006	2.642	2.460
Laplace Distribution $a = 1.5$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	1.448 (.003)	1.356 (.005)	1.307 (.007)
	1^{st} order	1.266	1.266	1.266
	2^{nd} order	1.456	1.361	1.314

Table 2. Variance approximations for L-estimates with $J = J_{a,F}$.

Table 2 is arranged as table 1. The first column gives the distribution for which estimates of location are being made as well as the parameter a of this L-estimate. The column labelled 'Variance approximation' gives a brief description of the values presented in each row. The last three columns are headed by the sample size for the estimates considered. The first row of each section presents the simulation size used for the Monte Carlo approximation of n times the variance given in the second row. The standard error of this approximation is given in parentheses. The third and fourth rows give the first and second order approximations, respectively, of n times the variance obtained from proposition 4.2.6.

When one wishes to compare variances of estimators from a class such as that described by the M-estimates corresponding to (5.2.1) or the L-estimates of (5.2.5) the expansions presented here can be particularly useful. Using simulation results to compare more than a few members of such a class would require an exorbitant amount of computer time. The variance approximations of propositions 4.2.6 and 4.3.3 take little computing time in comparison. Figures 3 and 4 present variance approximations for M-estimates corresponding to (5.2.1) and L-estimates corresponding to (5.2.5) when the underlying distribution is the Cauchy distribution given in (5.2.3).

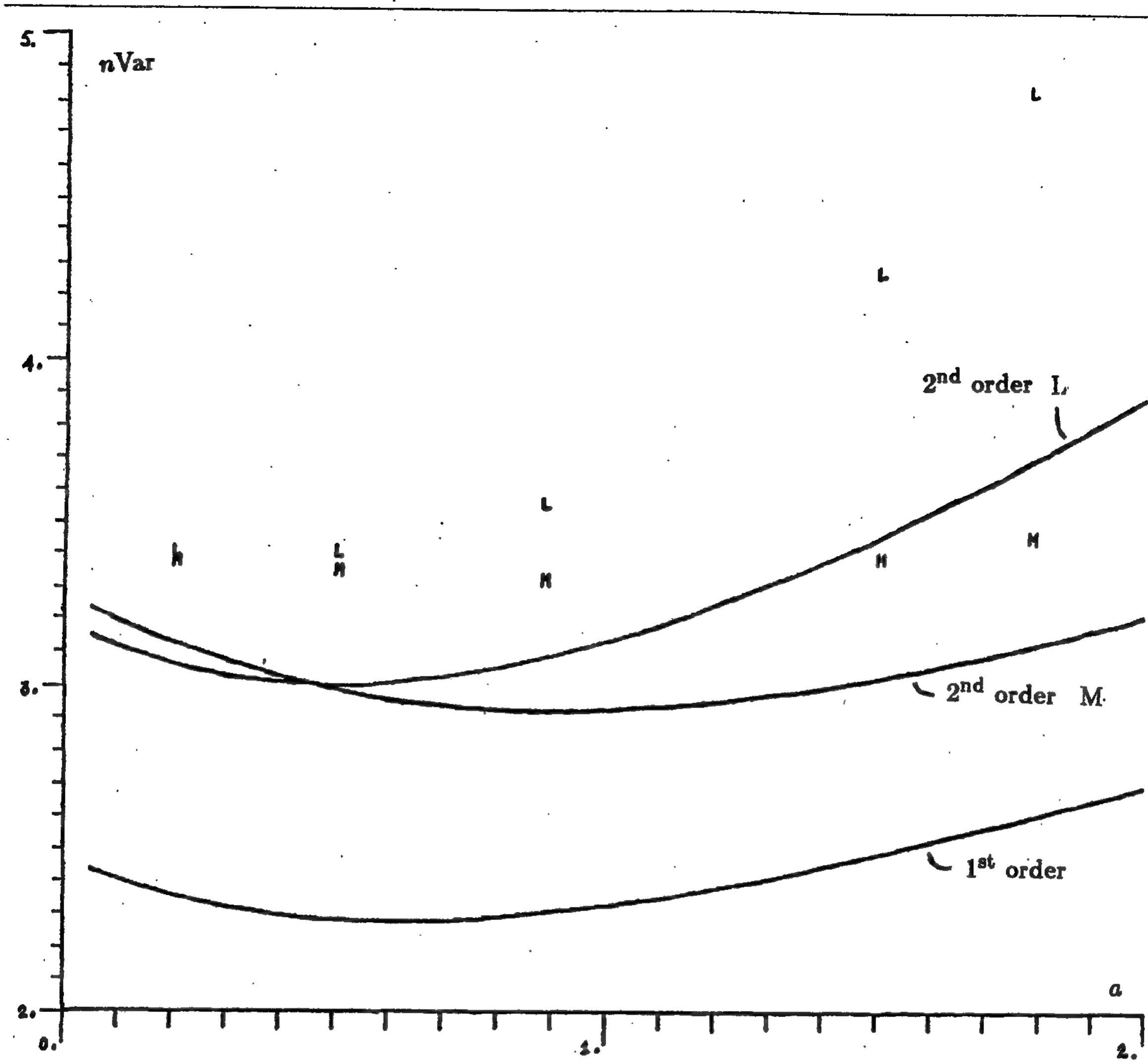

Figure 3. Variance approximations for L- and M-estimates; n=10.

Figures 3 and 4 present variance approximations for M-estimates corresponding to (5.2.1) and L-estimates corresponding to (5.2.5) when the underlying distribution is the Cauchy distribution given in (5.2.3). The lower line gives the first order variance approximations for both estimates as a function of a. The middle and upper lines give the second order approximation for M-estimates and L-estimates, respectively. The M's and L's plotted are simulation approximations for M- and L-estimates, respectively. The standard errors of the simulation results are about .03 except for the L-estimates corresponding to $n = 10$ and the two largest values of a where the standard errors are .08 and .15, respectively. Note that the scales on the y-axes of the two graphs differ.

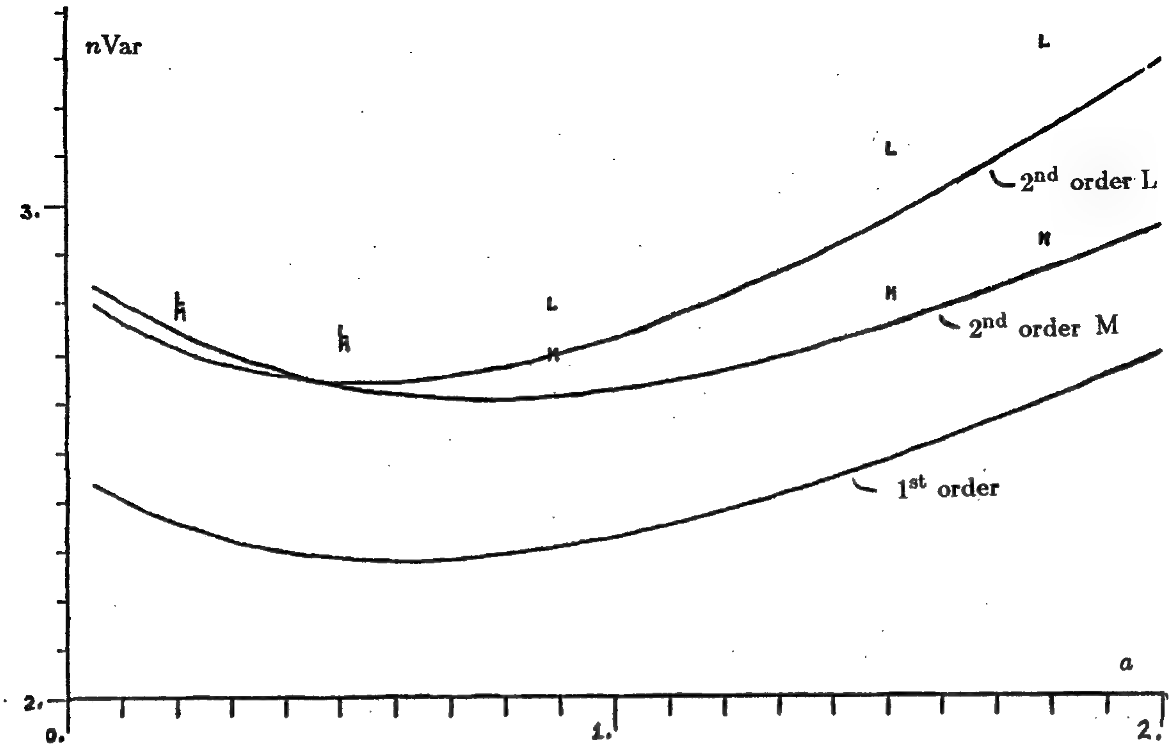

Figure 4. Variance approximations for L- and M-estimates; n=20.

There are several things worth noting in these figures. First, the error in the second order approximation is about one fourth as large for $n = 20$ as for $n = 10$. The second order approximation is much improved over the first order approximation. In choosing an optimum value of a it is clearly important to consider the second order approximation rather than just the first order approximation. Although the second order curves cross, it appears that for $n = 10$ or 20 there is no value of a for which the L-estimate is better than the M-estimate. Finally, for L-estimates the approximation gets worse as a increases, but for M-estimates this is not the case. This is undoubtedly related to the fact that for $a \geq F^{-1}(1 - 2/n)$ the second moment of the L-estimate is infinite.

Hodges and Lehmann (1970) discuss the comparison of variances of estimators which have the same first order efficiency. Assuming

$$n\mathrm{Var}_F(T^{(1)}(F_n)) = \sigma_1^2 + (1/n)\sigma_{21}^2 + o(1/n)$$

and

$$n\mathrm{Var}_F(T^{(2)}(F_n)) = \sigma_1^2 + (1/n)\sigma_{22}^2 + o(1/n)$$

where $\sigma_{22} > \sigma_{21}$, they suggest a measure called deficiency (or asymptotic expected deficiency) defined by

$$d = \frac{\sigma_{22}^2 - \sigma_{21}^2}{\sigma_1^2}. \tag{5.2.6}$$

They note that as n becomes large, the number of additional observations d_n needed to make

$$n\mathrm{Var}_F(T^{(2)}(F_{n+d_n})) = \mathrm{Var}_F(T^{(1)}(F_n)) + o(1/n^2)$$

tends to d. As an example we let $T^{(1)}(F_n)$ be an M-estimate with $\psi = \psi_a$ and let $T^{(2)}(F_n)$ be an L-estimate with $J = J_{a,F}$ where F is the Cauchy distribution of (5.2.3) and $a = 1.8$. The first coefficient of the variance approximation is 2.6 and the second coefficients are 5.2 (M) and 11.0 (L). This and (5.2.6) suggest that approximately $d = (11.0 - 5.2)/2.6 = 2.2$ additional observations are needed to get the same variance for the L-estimate as for the M-estimate as n becomes large. Table 3 indicates that for this example n must be moderately large before the variances of $T^{(1)}(F_n)$ and $T^{(2)}(F_{n+[\![d]\!]})$ become very close. Note that even the lines labelled '2$^{\text{nd}}$ order' are not that close. For most of the other examples considered in this section it appears that the asymptotic expected deficiency is less than 1.

	Variance Approximation			
M-estimate	Sample size	$n = 8$	$n = 18$	$n = 38$
	Simulation size	8,000	4,000	2,000
	Simulation (Std. Err.)	3.831 (.089)	2.974 (.055)	2.794 (.062)
	1^{st} order	2.608	2.608	2.608
	2^{nd} order	3.262	2.899	2.746
L-estimate	Sample size	$n = 10$	$n = 20$	$n = 40$
	Simulation size	50,000	25,000	12,500
	Simulation (Std. Err.)	4.857 (.146)	3.339(.029)	2.918 (.027)
	1^{st} order	2.608	2.608	2.608
	2^{nd} order	3.705	3.156	2.882

Table 3. Deficiency example: Cauchy distribution, $a = 1.8$

Table 3 is arranged somewhat differently than tables 1 and 2. The M-estimate considered has the ψ function given by (5.2.1). The L-estimate considered has the J function given by (5.2.5). The column labelled 'Variance approximation' gives a brief description of the values presented in each row. The first row of each section presents the sample size for the estimate of interest. The second row of each section presents the simulation size used for the Monte Carlo approximation of n times the variance given in the third row. The standard error of this approximation is given in parentheses. The fourth and fifth rows give the first and second order approximations, respectively, of n times the variance obtained from propositions 4.2.6 and 4.3.3.

F, a	Variance Approximation	$n = 10$	$n = 20$	$n = 40$
Normal Distribution $a = 1$	Simulation size	40,000	20,000	10,000
	Simulation (Std. Err.)	1.164 (.001)	1.183 (.002)	1.192 (.003)
	1^{st} order	1.208	1.208	1.208
	2^{nd} order	1.182	1.195	1.202
Cauchy Distribution $a = .6$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	3.411 (.016)	2.744 (.012)	2.485 (.012)
	1^{st} order	2.278	2.278	2.278
	2^{nd} order	3.006	2.642	2.460
Laplace Distribution $a = 1.5$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	1.476 (.003)	1.373 (.005)	1.316 (.007)
	1^{st} order	1.266	1.266	1.266
	2^{nd} order	1.456	1.361	1.314

Table 4. Variance approximations for L-estimates with $J(i/(n+1))$ coefficients; $J = J_{a,F}$.

Table 4 is the same as table 2 except that the L-estimates simulated use coefficients $J(i/(n+1))$ instead of $\int_{(i-1)/n}^{i/n} J(u)du$, $i = 1, 2, \ldots, n$. We have normalized these coefficients so that they sum to one. The table is arranged as tables 1 and 2 are. The first column gives the distribution for which estimates of location are being made as well as the parameter a of this L-estimate. The column labelled 'Variance approximation' gives a brief description of the values presented in each row. The last three columns are headed by the sample size for the estimates considered. The first row of each section presents the simulation size used for the Monte Carlo approximation of n times the variance given in the second row. The standard error of this approximation is given in parentheses. The third and fourth rows give the first and second order approximations, respectively, of n times the variance obtained from proposition 4.2.6.

The first thing to note in table 4 is that, as before, all expansions except for the Cauchy distribution with $n = 10$ appear quite good. The error of the first order approximation always appears to halve as n doubles. The error of the second order approximation for the normal distribution goes down by a factor of two as n quadruples. For the Cauchy distribution however, this error appears to quarter as n doubles, as before. Because of the standard error of the Monte Carlo approximation, the error behavior for the Laplace example is unclear. Whether or not the second order approximation is a correct one when coefficients for an L-estimate are computed in this fashion is not readily apparent from these examples. We have not attempted to justify this

expansion theoretically.

§5.3 Nonparametric variance estimates.

Since it is rarely the case that the underlying distribution function is known, we wish to give a brief example indicating that these expansions may be useful in approximating variances if we substitute the empirical distribution function in our formulas.

The variance of a functional $T(F_n)$ can be considered a functional of the underlying distribution function F, namely

$$\sigma^2(n,F) \equiv \mathrm{Var}_F(T(F_n)). \tag{5.3.1}$$

In chapter 4 we have given formulas approximating $\sigma^2(n,F)$ by an expression of the form

$$\sigma^2(n,F) = \frac{1}{n}\sigma_1^2(F) + \frac{1}{n^2}\sigma_2^2(F) + o(1/n^2). \tag{5.3.2}$$

We briefly consider the nonparametric variance approximations

$$n\sigma^2(n,F) \approx \sigma_1^2(F_n) \tag{5.3.3}$$

and

$$n\sigma^2(n,F) \approx \sigma_1^2(F_n) + \frac{1}{n}\sigma_2^2(F_n). \tag{5.3.4}$$

These approximations are 'delta method' approximations and can also be considered as first and second order approximations of the bootstrap estimate of variance, namely $\sigma^2(n,F_n)$.

Since the standard deviation of $\sigma_i^2(F_n)$, $i = 1,2,\ldots$ is, in general, $O(1/\sqrt{n})$ one might expect that the second (or any higher order term) of (5.3.2) would be useless. The reason we have considered this term is that it provides a second order approximation of the bootstrap which Efron (1981) has noted can be a better approximation than the first order delta method (note: Efron usually refers to the first order delta method as the infinitesimal jackknife).

As an example of the type of calculation to be done we recall the first order variance approximation of (4.2.7) and note that to obtain the right hand side of (5.3.3) we compute

$$\begin{aligned}\int\int\Big(F_n(\min(x_1,x_2)) - F_n(x_1)F_n(x_2)\Big)J(F_n(x_1))J(F_n(x_2))dx_1dx_2 \\ = \sum_{i=1}^{n-1}(X_{i+1:n} - X_{i:n})^2(i/n)(1-i/n)\big(J(i/n)\big)^2 \\ + 2\sum_{i=1}^{n-2}(X_{i+1:n} - X_{i:n})(i/n)J(i/n)\sum_{j=i+1}^{n-1}(X_{j+1:n} - X_{j:n})(1-j/n)J(j/n).\end{aligned}$$

F, a	Variance Approximation	$n = 10$	$n = 20$	$n = 40$
Normal Distribution $a = 1$	Simulation size	40,000	20,000	10,000
	Simulation (Std. Err.)	1.181 (.001)	1.195 (.002)	1.198 (.003)
	1^{st} order	1.208	1.208	1.208
	2^{nd} order	1.182	1.195	1.202
	1^{st} delta (Std. Dev.)	1.377 (1.008)	1.298 (.675)	1.260 (.466)
	2^{nd} delta (Std. Dev.)	1.364 (.779)	1.216 (.559)	1.242 (.426)
	1^{st} delta CV	.74	.35	.17
	2^{nd} delta CV	.44	.26	.15
	CV bound	.22	.11	.05
Cauchy Distribution $a = .6$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	3.404 (.016)	2.743 (.012)	2.484 (.012)
	1^{st} order	2.278	2.278	2.278
	2^{nd} order	3.006	2.642	2.460
	1^{st} delta (Std. Dev.)	5.553 (13.910)	3.342 (3.437)	2.759 (1.777)
	2^{nd} delta (Std. Dev.)	-.057 (30.913)	4.114 (5.394)	2.466 (1.496)
Laplace Distribution $a = 1.5$	Simulation size	160,000	80,000	40,000
	Simulation (Std. Err.)	1.448 (.003)	1.356 (.005)	1.307 (.007)
	1^{st} order	1.266	1.266	1.266
	2^{nd} order	1.456	1.361	1.314
	1^{st} delta (Std. Dev.)	1.765 (1.381)	1.510 (.835)	1.387 (.541)
	2^{nd} delta (Std. Dev.)	1.479 (1.135)	1.505 (.791)	1.431 (.535)

Table 5. Nonparametric variance approximations for L-estimates with $J = J_{a,F}$.

Table 5 is an expanded version of table 2. The first column gives the distribution for which estimates of location are being made as well as the parameter a of this L-estimate. The column labelled 'Variance approximation' gives a brief description of the values presented in each row. The last three columns are headed by the sample size for the estimates considered. The first row of each section presents the simulation size used for the Monte Carlo approximation of n times the variance given in the second row. The standard error of this approximation is given in parentheses. The third and fourth rows give the first and second order approximations, respectively, of n times the variance obtained from proposition 4.3.3. The rows labeled '1^{st} delta' give the average value of $\sigma_1^2(F_n)$ as in (5.3.3) from the simulation. The rows labeled '2^{nd} delta' give the average value of the right hand side of (5.3.4). Included in these rows are the estimated standard deviations of these estimators. It can be argued that a lower bound for the coefficient of variation of any location and scale invariant scale estimate for normal observations is that of s^2, namely $2/(n-1)$. This value is given in the row labeled 'CV bound'. The estimated coefficients of variation for the normal case are labelled '1^{st}

delta CV' and '2^{nd} delta CV'.

For the Cauchy distribution these approximations appear to be quite poor. For the normal and Laplace distributions the approximations appear reasonably good with the second order approximation appearing to have both lower bias and lower variance than the first order approximation; the only exception to this is that the bias is higher for the second order delta method for the Laplace distribution with $n = 40$.

It appears that further study of second and higher order delta method approximations might be worthwhile. From table 5 it appears that an important part of such a study would be to formulate estimates of the variation of such approximations.

§5.4 Quantiles and trimmed means.

In this section we present a pair of 'non-applications' of the theory we have developed. One might hope that the moment expansion of the limit of a set of estimators is the same as the limit of the moment expansions of these estimators. If this were true one could obtain variance approximations for trimmed means, quantiles, and other estimators which do not satisfy the assumptions of the moment convergence propositions that we have given. We give examples where the limit of the variance approximations of estimators is not equal to the variance expansion of the limit of the estimates. It will be seen, however, that we may improve a variance approximation substantially by using an (incorrect) expansion developed by taking the limit of expansions.

There are many simple functionals which are not Fréchet differentiable. Quantiles and linear combinations of quantiles are among these. We show this for a particular case as an example of the type of problem that may arise with a 'well behaved' functional.

Let $T(F) = \inf\{q : F(q) \geq c\}$ where $c \in (0,1)$, $F \in \mathcal{D}$. Note that $T(F) = \infty$ if the defining set is empty. Let $F(x) = x$ on $[0,1]$. We shall show that T is not Fréchet differentiable at F. For $\lambda > 0, y \in \Re$, let $F_{\lambda,y} = F + \lambda(\delta_y - F)$. For y fixed and λ sufficiently small we have for $y \neq c$ $T(F_{\lambda,y}) = (c - \lambda\delta_y(c))/(1-\lambda)$. The Gâteaux differential of T at F in the direction of $\delta_y - F$ is by definition

$$\lim_{\lambda\to 0} \frac{T(F_{\lambda,y}) - T(F)}{\lambda} = \lim_{\lambda\to 0} \frac{\lambda(c - \delta_y(c))}{\lambda(1-\lambda)} = c - \delta_y(c).$$

If the Fréchet differential exists, clearly it must be equal to the Gâteaux differential. Thus to show that T is not Fréchet differentiable at F it suffices to show that for some $\epsilon > 0$ and any $\delta > 0$ there exists $F_{\lambda,y}$ such that $\| F_{\lambda,y} - F \|_\infty < \delta$ and $|T(F_{\lambda,y}) - T(F) - \lambda(c - \delta_y(c))|$ $> \epsilon \| F_{\lambda,y} - F \|_\infty$. Fix $\epsilon \in (0,c)$. Let $\delta \in (0, 1-c)$ be arbitrary. Let $\lambda < \delta$. If $0 < y < 1$

then $\| F_{\lambda,y} - F \|_\infty = y + \lambda(1-y) - y = \lambda(1-y) < \delta$. Suppose $c < y < c/(1-\lambda)$. Then $T(F_{\lambda,y}) = y$ and

$$T(F_{\lambda,y}) - T(F) - \lambda(c - \delta_y(c)) = y - c - \lambda c.$$

Since for each $\lambda > 0$

$$\inf_{c<y<c/(1-\lambda)} \frac{y-c-\lambda c}{\lambda} = -c < -\epsilon$$

it follows that T is not Fréchet differentiable at F.

Since a quantile is not Fréchet differentiable we cannot apply theorem 3.4.1 to approximate moments. We now try to find a second order variance approximation for the median by taking the limit of expansions of M-estimates which approach the median. Let ψ be a continuous non-decreasing, non-constant, odd function on $\Re$ such that for $x \geq 1$, $\psi(x) = 1/2$. Assume also that ψ' and ψ'' exist everywhere and that ψ''' exists and is bounded everywhere except possibly at ± 1. For any positive a let $\psi_a(x) \equiv \psi(x/a)$. From proposition 4.3.3 we know that if F is symmetric and differentiable at a and satisfies the necessary tail conditions then the variance of the M-estimate corresponding to ψ_a for a sample of size n from F may be written as in (4.3.10). Letting $a \to 0$ it can be shown that if F is symmetric and three times differentiable at 0 then this is

$$\left(\frac{1}{n} - \frac{5}{3n^2}\right)\frac{1}{4(f(0))^2} - \frac{f''(0)}{16n^2(f(0))^5} + o(n^{-2}) \tag{5.4.1}$$

provided $f(0) > 0$. David (1980), p. 81 gives an expansion for the variance of a quantile. In the case of the median with F symmetric the formula he presents reduces to

$$\left(\frac{1}{n} - \frac{2}{n^2}\right)\frac{1}{4(f(0))^2} - \frac{f''(0)}{16n^2(f(0))^5} + o(n^{-2}). \tag{5.4.2}$$

In this case as $a \to 0$ the estimators corresponding to ψ_a converge to the median. Whereas the moment expansions of (5.4.1) and (5.4.2) are not the same, they are very similar. It appears that there is an 'extra term' when we do the calculation to obtain (5.4.1).

The trimmed mean is another case where we might try to get a variance approximation by taking the limit of variance approximations of statistics which approach the trimmed mean. In this case we will have three times Fréchet differentiable functionals approaching a trimmed mean which is one time Fréchet differentiable. Thus we have a 'smoother' situation than with the median above where the limiting functional was not Fréchet differentiable. The approximation obtained in this fashion appears likely, once again, to be incorrect.

The α-trimmed mean is an L-statistic with weight function

$$J(u) = \begin{cases} \frac{1}{1-2\alpha}, & \text{if } \alpha \leq u \leq 1-\alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{5.4.3}$$

Because J is not differentiable at α and $1-\alpha$ we may not apply proposition 4.2.6 to obtain second order variance approximations. A smoothed version of this weight function is given by

$$J(u) = \begin{cases} \frac{1}{1-2\alpha}, & \text{if } \alpha+\epsilon \le u \le 1-\alpha-\epsilon, \\ \frac{1}{1-2\alpha} w\left(\frac{u-\alpha}{\epsilon}\right), & \text{if } \alpha-\epsilon \le u \le \alpha+\epsilon, \\ \frac{1}{1-2\alpha} w\left(\frac{1-\alpha-u}{\epsilon}\right), & \text{if } 1-\alpha-\epsilon \le u \le 1-\alpha+\epsilon, \\ 0, & \text{otherwise,} \end{cases} \tag{5.4.4}$$

where w is defined on $[-1,1]$ and has the following properties: 1) $w(-1)=0$, $w(1)=1$; 2) w is three times continuously differentiable on $[-1,1]$ with $w'(-1)=w'(1)=0$; 3) w is symmetric about 0, *i.e.* $w(u)=1-w(-u)$.

Property 3 implies $\int_{-1}^{1} w(x)dx = 1$. This implies that if J is defined as in (5.4.4) then $\int_0^1 J(u)du = 1$. The L-statistic corresponding to this weight function will be referred to as an ϵ-smoothed, α-trimmed mean. For any ϵ-smoothed, α-trimmed mean we have J and $J^{(1)}$ continuous and bounded on $[0,1]$, and $J^{(2)}$ continuous and bounded except possibly at $\alpha \pm \epsilon$ and $1-\alpha\pm\epsilon$. It follows that if

$$0 < \liminf_{x\to\infty} \frac{-\log(1-F(x))}{\log x}, \quad 0 < \liminf_{x\to\infty} \frac{-\log F(-x)}{\log x} \tag{5.4.5}$$

and either F^{-1} is continuous at $\alpha\pm\epsilon$ and $1-\alpha\pm\epsilon$ or $w''(-1)=w''(1)=0$ then we may apply proposition 4.2.6 to obtain bias and first and second order variance approximations.

We consider the case where F is symmetric about zero, is two times differentiable at $F^{-1}(\alpha)$, and satisfies (5.4.5). We let $\delta > 0$ be such that $f(x) = \frac{d}{dx}F(x)$ exists in $(F^{-1}(\alpha-\delta), F^{-1}(\alpha+\delta))$. For any $\epsilon \in (0,\delta)$ it follows that F^{-1} is continuous at $\alpha\pm\epsilon$, $1-\alpha\pm\epsilon$ and we may apply proposition 4.2.6. For the first order variance term we will apply (3.5.9) rather than (4.2.7). Note that $T(F)=0$. From (4.2.4) and definition 3.3.1 we can show that if $z=\delta_x - F$ then

$$T_1(F;z) = \begin{cases} x/(1-2\alpha), & \text{if } |x| \le F^{-1}(1-\alpha-\epsilon), \\ \frac{\operatorname{sgn} x}{1-2\alpha}\left(F^{-1}(1-\alpha-\epsilon) + \int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dy\right), & \text{if } |x| \ge F^{-1}(1-\alpha+\epsilon), \\ \frac{\operatorname{sgn} x}{1-2\alpha}\left(F^{-1}(1-\alpha-\epsilon) + \int_{-|x|}^{F^{-1}(\alpha+\epsilon)} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dy\right), & \text{otherwise.} \end{cases} \tag{5.4.6}$$

This implies that the first order variance approximation has coefficient

$$
\begin{aligned}
\mathrm{E}[(T_1(F;Z_1))^2] &= \frac{2}{(1-2\alpha)^2}\int_0^{F^{-1}(1-\alpha-\epsilon)} x^2 dF(x) + \frac{2(\alpha+\epsilon)}{(1-2\alpha)^2}(F^{-1}(1-\alpha-\epsilon))^2 \\
&+ \frac{2(\alpha-\epsilon)}{(1-2\alpha)^2}\left(\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dy\right)^2 \\
&+ \frac{4(\alpha-\epsilon)F^{-1}(1-\alpha-\epsilon)}{(1-2\alpha)^2}\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dy \\
&+ \frac{2}{(1-2\alpha)^2}\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)}\left(\int_{F^{-1}(\alpha-\epsilon)}^{x} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dy\right)^2 dF(x) \\
&+ \frac{4F^{-1}(1-\alpha-\epsilon)}{(1-2\alpha)^2}\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)}\int_{y}^{F^{-1}(\alpha+\epsilon)} w\left(\frac{F(y)-\alpha}{\epsilon}\right)dydF(x) \\
&= \frac{2}{(1-2\alpha)^2}\int_0^{F^{-1}(1-\alpha)} x^2 dF(x) + \frac{2\alpha}{(1-2\alpha)^2}(F^{-1}(1-\alpha))^2 + O(\epsilon).
\end{aligned} \tag{5.4.7}
$$

Applying (4.2.9), (5.4.4) and the properties of w, a series of straightforward calculations shows that the following is the coefficient of $1/n^2$ in the second order variance approximation:

$$
\begin{aligned}
&\frac{2}{(1-2\alpha)^2\epsilon}\Bigg\{\left(F^{-1}(1-\alpha-\epsilon)+\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} F(x)w\left(\frac{F(x)-\alpha}{\epsilon}\right)dx\right) \\
&\qquad\qquad \cdot\left(\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} F(x)(1-2F(x))w'\left(\frac{F(x)-\alpha}{\epsilon}\right)dx\right) \\
&+\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} F(x_1)w\left(\frac{F(x_1)-\alpha}{\epsilon}\right)\int_{x_1}^{F^{-1}(\alpha+\epsilon)}(1-F(x_2))(1-2F(x_2))w'\left(\frac{F(x_2)-\alpha}{\epsilon}\right)dx_2dx_1 \\
&-\frac{1}{\epsilon}\left(\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)}(F(x))^2w'\left(\frac{F(x)-\alpha}{\epsilon}\right)dx\right)^2 \\
&+\frac{2}{\epsilon}\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)}(F(x_1))^2w'\left(\frac{F(x_1)-\alpha}{\epsilon}\right)\int_{x_1}^{F^{-1}(\alpha+\epsilon)}(1-F(x_2))^2w'\left(\frac{F(x_2)-\alpha}{\epsilon}\right)dx_2dx_1 \\
&+\frac{1}{\epsilon}\left(F^{-1}(1-\alpha-\epsilon)+\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} F(x)w\left(\frac{F(x)-\alpha}{\epsilon}\right)dx\right) \\
&\qquad\qquad \cdot\left(\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)}(F(x))^2(1-F(x))w''\left(\frac{F(x)-\alpha}{\epsilon}\right)dx\right) \\
&+\frac{2}{\epsilon}\int_{F^{-1}(\alpha-\epsilon)}^{F^{-1}(\alpha+\epsilon)} F(x_1)w\left(\frac{F(x_1)-\alpha}{\epsilon}\right)\int_{x_1}^{F^{-1}(\alpha+\epsilon)} F(x_2)(1-F(x_2))^2w''\left(\frac{F(x_2)-\alpha}{\epsilon}\right)dx_2dx_1\Bigg\}.
\end{aligned} \tag{5.4.8}
$$

Further straightforward calculations show that if F is two times differentiable at $F^{-1}(\alpha)$ then this is equal to

$$
\frac{2\alpha(1-\alpha)F^{-1}(1-\alpha)}{(1-2\alpha)^2 f(F^{-1}(\alpha))}\left(\alpha\frac{f'(F^{-1}(\alpha))}{(f(F^{-1}(\alpha)))^2}-1\right)+\frac{\alpha^2}{(1-2\alpha)(f(F^{-1}(\alpha)))^2}+o(1) \text{ as } \epsilon\to 0. \tag{5.4.9}
$$

Equations (5.4.7) and (5.4.9) suggest the following approximation for the α-trimmed mean when F is symmetric and two times differentiable at $F^{-1}(\alpha)$ and (5.4.5) holds:

$$\frac{1}{n}\frac{2}{(1-2\alpha)^2}\left(\int_0^{F^{-1}(1-\alpha)} x^2 dF(x) + \alpha(F^{-1}(1-\alpha))^2\right)$$
$$+\frac{1}{n^2}\left(\frac{2\alpha(1-\alpha)F^{-1}(1-\alpha)}{(1-2\alpha)^2 f(F^{-1}(\alpha))}\left(\alpha\frac{f'(F^{-1}(\alpha))}{(f(F^{-1}(\alpha)))^2}-1\right)+\frac{\alpha^2}{(1-2\alpha)(f(F^{-1}(\alpha)))^2}\right). \tag{5.4.10}$$

Distribution	%trim	n	Exact	1[st] order (Err.)	2[nd] order (Err.)	Monte Carlo Approx. (Err.)
Normal	10	5	1.019	1.060 (.041)	1.031(.012)	1.020(.001)
		10	1.053	1.060 (.007)	1.046 (-.007)	1.048 (-.005)
		20	1.055	1.060 (.005)	1.053 (-.002)	1.056 (.001)
	25	5	1.145	1.195 (.050)	1.144 (-.001)	1.156 (.011)
		10	1.164	1.195 (.031)	1.170 (.006)	1.148 (-.016)
		20	1.186	1.195 (.009)	1.182 (-.004)	1.199 (.013)
Laplace	10	5	1.758	1.494 (-.264)	1.825 (.067)	
		10	1.617	1.494 (-.123)	1.659 (.042)	
		20	1.556	1.494 (-.062)	1.577 (.021)	1.60 (.04)
	25	5	1.599	1.227 (-.372)	1.766 (.167)	
		10	1.424	1.227 (-.197)	1.497 (.073)	
		20	1.228	1.227 (-.001)	1.362 (.134)	1.33 (.10)
Cauchy	10	20	8.282	4.771 (-3.511)	6.78 (-1.50)	7.3 (-1.0)
		40		4.771	5.77	5.40
	25	10	4.498	2.546 (-1.952)	3.58 (-.92)	4.6 (.1)
		20	3.182	2.546 (-.636)	3.06(-.12)	3.1(-.1)
		40		2.546	2.80	2.61

Table 6. Variance approximations for trimmed means.

Table 6 contains exact values and various approximations for n times the variance of various trimmed means. We have used the approximation in (5.4.7) and (5.4.10) to compute the approximations of n times the variance given in the columns labelled '1[st] order' and '2[nd] order', respectively. Some of the exact numbers were found in Gastwirth and Cohen (1970). Other exact variances were computed using tables of variances and covariances of order statistics. These tables are given by Sarhan and Greenberg (1964) (normal), Govindarajulu (1966) (Laplace), and Barnett (1968) (Cauchy). In the last column of the table are Monte Carlo approximations of the variances of trimmed means which can be found in Andrews *et.al.* (1972). We do not consider the trimmed mean for the Cauchy distribution with 10% trim and $n = 10$ as the true variance is infinite.

Looking at these numbers carefully suggests that the second order term of (5.4.10) is not

correct as the difference of the columns labelled 'Exact' and 'Approx.' often decreases only by a factor of about two as n doubles. Note also, however, that even these apparently incorrect second order approximations can be a great improvement over first order expansions. The Monte Carlo approximations given used variance reduction techniques and simulation sizes of 640 to 1000. The error of these approximations is comparable to the error of 2^{nd} order approximation given. We have not attempted to rigorously derive a correct version of the second order variance approximation of the trimmed mean. Because of the widespread interest in trimmed means such a derivation might be worthwhile.

References

1. Anderson, K., Sobel, M., and Uppuluri, V. R. R. (1982). Quota fulfillment times. To appear in *The Canadian Journal of Statistics.*

2. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location.* Princeton University Press.

3. Barnett, V.D. (1966). Order statistics estimators of the location of the Cauchy distribution. *Journal of the American Statistical Association* **61**, 1205-1218. Correction (1968). **63**, 383-385.

4. Bickel, P. J. (1967). Some contributions to the theory of order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**, 575-591. University of California Press.

5. Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables.* Wiley, New York.

6. Boos, D. B. (1979). A differential for L-statistics. *Annals of Statistics* **7**, 955-959.

7. Boos, D. B., and Serfling, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions with application to M-estimates. *Annals of Statistics* **8**, 618-624.

8. Box, G. E. P. and Muller, M. A. (1959). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics* **29**, 610-613.

9. Breiman, L. (1968). *Probability.* Addison-Wesley, Reading, Massachusetts.

10. Chung, K. L. (1974). *A Course in Probability Theory.* Academic Press, London.

11. David, F. N. and Johnson, N. L. (1954). Statistical treatment of censored data. I. Fundamental formulae. *Biometrika* **41**, 223-240.

12. David, H. A. (1980). *Order Statistics.* Wiley, New York.

13. de Haan, L. (1970). *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes.* Mathematical Centre Tracts **32**. Mathematical Centre, Amsterdam.

14. Dvoretsky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic mimimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics* **27**, 642-669.

15. Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68**, 589-600.

16. Eynon, B. (1982). Ph.D. dissertation, Stanford University. In preparation.

17. Feller, W. (1971). *An Introduction to Probability Theory and its Applications* **2**. Wiley, New York.

18. Gastwirth, J.L. and Cohen, M.L. (1970) Small sample behavior of robust linear estimators of location. *Journal of the American Statistical Association* **65**, 946-973.

19. Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and applications. *Annals of Mathematical Statistics* **42**, 1957-1961.

20. Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population. *Journal of the American Statistical Association* **61**, 248-258.

21. Hardy, G. H. (1952). *A Course of Pure Mathematics*, 10^{th} edition. Cambridge University Press, New York.

22. Hodges, J. L. and Lehmann, L. (1970). Deficiency. *Annals of Mathematical Statistics* **41**, 783-801.

23. Holst, Lars (1981). On sequential occupancy problems. *Journal of Applied Probability* **18**, 435-442.

24. Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

25. Johns, M. V., Jr (1979). Robust Pitman-like estimators. *Robustness in Statistics*, 49-59, Academic Press.

26. Knuth, D. E. (1969). *The Art of Computer Programming,* **2** *Seminumerical Algorithms.* Addison-Wesley, Reading, Massachusetts.

27. Mason, D. M. (1981). Asymptotic normality of linear combinations of order statistics with a smooth score function. *Annals of Statistics* **9**, 889-904.

28. Parzen, E. (1979). Nonparametric statistical data modelling. *Journal of the American Statistical Association* **74**, 105-131.

29. Pickands, J. III (1968). Moment convergence of sample extremes. *Annals of Mathematical Statistics* **39**, 881-889.

30. Reeds, J. A. III (1976). *On the definition of von Mises functionals.* Ph.D. dissertation, Harvard University.

31. Rényi, A. (1970). *Probability Theory.* North-Holland, Amsterdam.

32. Rudin, W. (1964). *Principles of Mathematical Analysis*, 2^{nd} edition. McGraw-Hill, New York.

33. Sarhan, A.E. and Greenberg, B.G. (Eds.) (1962). *Contributions to Order Statistics.* Wiley, New York.

34. Sen, P. K. (1959). On the moments of the sample quantiles. *Calcutta Statistical Association Bulletin* **9**, 1-19.

35. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

36. Simon, G. (1976). Computer simulation swindles, with applications to estimates of location and dispersion. *Applied Statistics* **3**, 266-274.

37. Smirnov, N. V. (1952). *Limit Distributions for the Terms of a Variational Series.* American Mathematical Society, Translation No. 67.

38. Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Annals of Statistics* **2**, 676-693. (1979). Correction note. *Annals of Statistics* **7**, 466.

39. von Mises, R. (1947). On the asymptotic distribution of differentiable functions. *The Annals of Mathematical Statistics* **18**, 309-348.

40. Wretman, J. (1978). A simple derivation of the asymptotic distribution of a sample quantile. *Scandanavian Journal of Statistics* **5**, 123-124.

Author Index

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* MOMENT EXPANSIONS FOR ROBUST STATISTICS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) KEAVEN MARTIN ANDERSON		8. CONTRACT OR GRANT NUMBER(s) DAAG29-79-C-0166
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPARTMENT OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIFORNIA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. ARMY RESEARCH OFFICE PO BOX 12211 RESEARCH TRIANGLE PARK, NC 27709		12. REPORT DATE MARCH 12, 1982 13. NUMBER OF PAGES 72
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

MOMENT EXPANSION, ASYMPTOTICS, FUNCTIONAL DIFFERENTIATION, L-ESTIMATES, M-ESTIMATES, QUANTILES, ROBUST ESTIMATION, LOCATION ESTIMATION, MOMENT GENERATING FUNCTIONS FOR QUANTILES.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

SEE OTHER SIDE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

Our objective is to give asymptotic expansions for moments of standardized statistics based on n independent, identically distributed random variables as $n \to \infty$. The basic premise is that a simple tail condition on the underlying distribution which implies the moments of a standardized quantile converge to the moments of an appropriate normal distribution is sufficient to assure the validity of asymptotic moment expansions for many statistics which are resistant to outliers.

The primary result we present gives sufficient conditions for the validity of moment approximations based on moments of Taylor's series approximations which are obtained by using functional differentiation. We apply the theory to some L- and M-estimates and present a Monte Carlo study to show that the approximations for the variance of statistics based on small to moderate sample sizes can be quite good.

Prior to studying the above general problem we consider the problem of the convergence of the moments of a standardized quantile to those of an appropriate normal distribution. Our proof of moment convergence requires fewer non-tail conditions on the underlying distribution than were used in previously published results. We also extend the result to show necessary and sufficient tail conditions on the underlying distribution for convergence of the moment generating function of a standardized quantile to that of a normal distribution.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)